

Development of a National seamless database
of topography and hydrologic derivatives

Sandra K. Franken
Dean J. Tyler
Kristine L. Verdin
Raytheon
EROS Data Center
Sioux Falls SD

Abstract

The recent completion of the [National Elevation Dataset](#) (NED) and the [National Hydrography Dataset](#) (NHD) has provided an avenue for nationwide development of topographically derived hydrologic data layers at a scale of 1:24,000. This multilayer dataset of hydrologic derivatives, entitled the [Elevation Derivatives for National Applications](#) (EDNA), is being developed by a consortium of participants, including the [U.S. Geological Survey](#), [National Weather Service](#), the [Environmental Protection Agency](#) and others. After the dataset is completed, terabytes of data will need to be stored, managed and distributed. To facilitate browse and, ultimately, delivery of the EDNA data, the consortium decided to manage the data layers with ArcSDE and provide browse and delivery of the data through the use of ArcIMS.

* Raytheon ITSS, USGS EROS Data Center, Sioux Falls, SD 57198 (Work performed under U.S. Geological Survey contract 1434-CR-97-CN-40274.) Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government

Introduction

Increasingly, many local, state, and federal agencies that have the mandate for management of water resources are finding that their needs are not being met by existing digital data sets. Current national coverage of digital data sets, such as, drainage basin boundaries and consistent elevation-derived parameters, do not exist or are not of a suitable scale or consistency to allow management of small or mid-size watersheds. This problem has become more significant as the management of water resources, both in terms of quantity and quality, is becoming more and more based on the watershed scale.

In 2000, the [U.S. Geological Survey](#) (USGS) [EROS Data Center](#) completed the first version of the periodically updated [National Elevation Dataset](#) (NED). Completion of this dataset has provided an avenue for nation-wide development of topographically derived hydrologic data layers at a scale of 1:24,000. The goal of this data development effort has been to develop, in a systematic and consistent fashion, many of the topographically-derived data layers used in modeling efforts. These derivative layers make up a multi-layered dataset, [entitled Elevation Derivatives for National Applications](#) (EDNA), being developed by a consortium of participants, including the [USGS National Mapping Division](#) (USGS/NMD), the [USGS Water Resources Division](#) (USGS/WRD), the [National Weather Service](#) (NWS), the [Environmental Protection Agency](#) (EPA), and others.

The impetus for the development of this dataset has been multi-faceted:

- The development of the EDNA dataset was responsive to the need for better drainage basin boundaries for the country. The new [Watershed Boundaries Dataset](#) (WBD) strives to identify the "best-available" watersheds boundaries available on a national level. Development of the EDNA derived Cataloging Unit, Watershed, and Subwatershed boundaries will be used to provide high-resolution boundaries for the WBD.
- The EDNA dataset will provide the capability of developing drainage basin boundaries above any point, and provide downstream linkages within the U.S. With this information available on a national scale, impacts of pollutant spills can be easily traced through the network, drainage areas above any point, not just terminal points of pre-defined watersheds, can be determined, and watershed units and stream segments downstream of a point-source discharge along with the location of the stream segment to which the point-source discharges can be easily be identified.
- Development of the EDNA dataset will serve to integrate two of the USGS' key national dataset, the [NED](#) and the [NHD](#). Enhancements to both datasets will be expected as the quality control procedures used in the development of the EDNA provide feedback to both NED and NHD. The NHD will be further enhanced by consistency with the EDNA. Elevation-derived streamline and basin parameters can be transferred onto the NHD following conflation with the EDNA. This will provide valuable attributes useful in model parameterization.

The development of EDNA required intense planning processes. The goal was to generate a seamless 30 meter product for the United States. While many of the tools used in the EDNA development have been implemented in standard Geographic Information System (GIS) software and used for small to moderate size applications, application of the tools to a problem of this size had not been attempted before. Due to the complexity of the project, three stages were necessary to process, store, manage, and distribute several terabytes of data. The

purpose of this paper is to discuss EDNA Stage I data layers, as well as, the procedures developed to load, manage, and distribute the data.

Materials and Methodology

EDNA Stage I Overview

If software and hardware capabilities had been up to the task, processing of the EDNA Stage I layers would have been done in one pass; processing the entire NED into hydrologic derivatives in one large piece. However, the NED data, which forms the basis for the EDNA dataset, is a very large Digital Elevation Model (DEM), requiring almost 60 GB of disk space for the conterminous United States alone. The size of the problem required a unique solution.

Utilizing the digital [Hydrologic Units of the United States](#) (Seaber, et-al, 1994), the EDNA Stage I data products were developed on a Cataloging Unit (CU) by Cataloging Unit basis. The Cataloging Units were chosen as the processing units, not only because they subdivided the NED into smaller, more manageable units, but because these units are “hydrologically based.” The boundaries of the Cataloging Unit more or less approximate hydrologic divides, which makes the task of assembling the pieces back into a seamless dataset somewhat easier.

Semi-Blind Pass Process and Editing

EDNA Stage I, or semi-blind pass processing, was designed to gather the required raw data, to organize that data, and to execute the preliminary processing

with the minimum amount of interaction between the processing system and the operator. EDNA Stage I is an ArcInfo process that results in the generation of 11 grids and 18 coverages. It consists of a set of arc macro languages (amls) that execute the required tasks in an almost fully automated procedure. For every Cataloging Unit in the U.S., the following procedures are undertaken to create the Stage I EDNA layers:

- The polygon(s) corresponding to the CU being processed is extracted from the 1:250,000 [Hydrologic Unit dataset](#) (Figure 1) (Seaber, et-al, 1994). This dataset is unprojected in geographic coordinates.
- The [extracted polygon](#) (Figure 2) is projected into the National Albers projection and is buffered by 25,000m. The buffer is necessary to mitigate any edge effects that may occur.
- The [NED data](#) (Figure 3) underlying the buffered polygon is identified, extracted from the seamless dataset and projected into the National Albers projection with a cell size of 30m.
- This projected and clipped DEM is filled using the standard ARC/INFO implementation of Jenson & Domingue (1988). In areas of the country with natural closed drainage, natural depressions can be maintained in the DEM. The derivative layers are developed, including flow direction, flow accumulation and slope.
- A first cut at derivation of [synthetic streamlines](#) (Figure 4) is made by thresholding the flow accumulation at 5,000 pixels (approximately 2 sq. mi.). Corresponding drainage areas for each reach in the synthetic streamline coverage are derived.
- The [synthetic streamline coverage is intersected](#) with the 1:250,000 Cataloging Unit boundary. A potential [seed point](#) is identified at every location that the streamlines intersect the Cataloging Unit boundary (Figure 4). At this point in the processing, user intervention is required to select the correct location of the seed points.
- The selected seed points are used to generate [a DEM-derived equivalent of the Cataloging Unit boundary](#) (Figure 5). This boundary is used as the template for clipping the raster data layers and merging them back together

into a [seamless unit](#) (Figure 6). Verification of the accuracy of the location of these seed points is essential to creating a seamless data set.

It is critical to approach the Stage I process with caution. Inaccurate placement of the selected seed point with respect to adjacent Cataloging Unit synthetic streamlines will lead to an accumulation of problems that will ripple outward affecting all watershed boundaries surrounding the one being processed. As a result, numerous Cataloging Units will need to be reprocessed to correct the areas affected by the initial error.

After completion of the automated process, a cursory quality review is conducted to identify gross inconsistencies between the Cataloging Unit boundary, or the drainage derived from the elevation data and the existing Hydrologic Unit boundary or NHD drainage pattern. The purpose of the review at this point in the procedure is to resolve major problems that, if left unresolved, have the potential to render further processing unacceptable. The review consists of a visual and statistical comparison of the derived Cataloging Unit boundary, and the derived drainage with the existing Hydrologic Unit boundary and NHD drainage pattern.

Before incorporating the Stage I data layers into the seamless data set, the data are checked to ensure that they are, in fact, seamless, without any gaps or overlaps. Additionally, the flow accumulation grid is adjusted to account for upstream drainage. The stream and drainage basins are attributed with a [Pfafstetter code](#) (Verdin & Verdin, 2000) to allow for rapid up and downstream tracing

(Figure 7). Following these procedures, the [data are loaded into SDE](#) for subsequent access, viewing, analyses and downloading portions of the dataset (Figure 8).

Description of EDNA Stage I Grids

The EDNA Stage I semi-blind pass process results in terabytes of data for each Cataloging Unit. The following grids are created and retained during the Stage I process. The grids are listed and described in the order in which they are created. Each of the grids covers an area equal to the extent of the Cataloging Unit plus the buffer zone of 25,000 meters.

- The original elevation data is the initial set of digital elevation data. These data are derived from the NED. The NED data are projected to an Albers Equal Area projection on NAD83. The cell size is 30 by 30 meters, and the vertical resolution is one centimeter. The grid is an integer grid.
- The [filled elevation grid](#) is created from the original elevation data by filling all of depressions, or sinks, in the original DEM (Figure 9). In areas of naturally occurring depression, capabilities are built into the Stage 1 algorithms to maintain natural depressions in the landscape.
- The [shaded relief grid](#) is created from original DEM grid using the “hillshade” command (Figure 10).
- The sinks grid defines the extent of each depression (sink). Each cell within each depression is assigned a unique value. Upland cells are assigned a value of “no data.”
- The flow direction grid is generated from the [filled DEM](#) using the “flowdirection” command (Figure 9). Each cell in the flow direction grid is assigned a code (value) that defines the direction water will flow from the cell. There are eight possible flow directions: east, southeast, south, southwest, west, northwest, north, and northeast.
- The flow accumulation grid is generated from the flow direction using the “flowaccumulation” command. Each cell in the flow accumulation grid is

assigned a value that is the sum of all of the cells that contribute flow to the cell.

- The [catchment basin grid](#) defines the area of all basins corresponding to a predefined set of “seedpoints” (Figure 11). In the Stage I processing methodology, a seed point is the first cell upstream of the confluence of two or more streams where the area drained by each stream exceeds a predetermined threshold value. The current threshold value used in the Stage I process is 5000 cells. At each confluence a catchment basin is determined for each stream forming the confluence. The area of each catchment basin is equal to the total upstream area minus the area of any upstream catchment basins. Each catchment basin is assigned a unique value, and all of the cells in the basin carry that value.
- The [slope grid](#) is generated from the filled DEM using the “slope” command with the “degree,” not “per cent” slope parameter (Figure 9).
- The threshold grid is a subset of the sinks grid. It contains only those sinks that exceed a predetermined area threshold. The threshold currently implemented is 1000 cells. This grid may be useful in selecting sinks to be retained for future processing steps.
- The locate grid is a “mask” derived from the threshold grid. It contains only the cells from the threshold grid that are located at the lowest point(s) in the sinks contained in the threshold grid. The value of the cells in the locates grid are -9999 or “no data.”
- The gen_huc_g grid is the DEM-derived equivalent of the Cataloging Unit boundary and is generated from the filled DEM using the “watershed” command and the selected outlet, and in some cases, inlet point(s) of the Cataloging Unit being processed. The Cataloging Unit numbers are assigned to all of the cells in the corresponding derived watershed. All remaining cells in the grid are assigned the “no data” value.

Description of EDNA Stage I Vector Coverages

- The [Cataloging Unit vector coverage](#) is extracted from the Hydrologic Unit dataset (Figure 2). This boundary delineates the extent of the 8-digit Hydrologic Unit being processed.
- The [buffered area vector coverage](#) delineates the extent of the 25,000 meter buffered zone encircling the 8-digit Hydrologic Unit (Figure 2).
- The metadata vector coverage maintains, via the INFO files, the record of information pertinent to the Hydrologic Unit being processed, information

about the DEM processing, and information from the NED metadata about the source of the elevation data.

- The [stream line vector coverage](#) is derived from the flow accumulation grid (Figure 4). A threshold, currently 5000 cells, is applied to the flow accumulation grid, and a mask is created. All cells having a value greater than 5000 are set to 1, and the remaining cells are set to NODATA. The mask grid is converted to the vector streamlines coverage.
- The [seed points point coverage](#) is a set of points established at the first cell upstream from the confluence of two or more streamlines in the stream_lines coverage (Figure 4). A seed point is the lowest point in a watershed, or in the case of a sink, it is the outlet point.
- The [sinks coverage](#) is the vector representation of the sinks grid (Figure 12).
- The [catchment coverage](#) is the vector representation of the catchment grid (Figure 11).
- The contour vector coverage is a set of contours derived from the [filled DEM grid](#) (Figure 9).
- The hydrologic data [NHD vector coverage](#) is extracted from the [NHD](#) (Figure 13). It contains the arcs defining the streamlines, water bodies, and centerlines. The NHD point data (NHDPT) coverage is extracted from the NHD. It defines the location of point features, such as, wells and springs. Both NHD and NHDPT have the same extent as the huc_poly coverage. The NHDDUU vector coverage is also extracted from the NHD. It contains the arcs that define the boundaries of the 1:100,000 scale quadrangles, and the arcs that define the extent of the Hydrologic Unit. This boundary is a close, but not always exact representation of the 8-digit Hydrologic Unit.
- The [point data, huc8 seeds](#), originally consists of a set of points that define the location of all intersections between the huc_poly coverage and the stream_lines coverage (Figure 4). Subsequently, the coverage is edited, and only the outlet point (pour point) and the inlet point(s) (catchment), if any exist, of the 8-digit Hydrologic Unit are retained. All other points are deleted. The edited coverage is the coverage that is retained during the processing.
- The [point coverage, locate_pts](#), is originally the point representation of the locates grid (Figure 14). The original locate_pts coverage is edited, and all points within the 8-digit Hydrologic Unit boundary are deleted unless there exists within the boundary a sink that is large enough to be retained. If there is such a sink, the point(s) in the locate_pts coverage that are associated with that sink are not deleted. If there is a large sink, the Stage I processing is

redone, and the point(s) remaining in the locate_pts coverage are used to prevent the filling of the target sinks.

- The NHD reach vector coverage is extracted from the NHD. It contains the arcs defining the NHD reaches and related metadata in the nhd_rch.aat file. The coverage has the same extent as the huc_poly coverage.
- The [generated hydrologic coverage, gen_huc_p](#), is the vector representation of the generated hydrologic grid (Figure 5).
- The clipped streams vector coverage is identical to the streamlines coverage except that it is clipped to the extent of the generated hydrologic coverage.
- The matched stream line vector coverage is spatially identical to the clipped streams coverage, but it has additional attributes in the “streams_match.aat” file that are extracted from the NHD files.
- The union vector coverage is a “union” of the huc_poly and the gen_huc_p coverages (Kelly, Description).

Problems and Solutions

The goal of the EDNA project was to create a seamless dataset derived from a hydrologically conditioned DEM. In theory, the Stage I process seemed simplistic at a smaller scale, such as the processing of one Hydrologic Unit. However, approaching Stage I at the magnitude of a national scale, where over [2,000 different watershed circumstances exist](#), warrants discussion of issues that needed to be resolved (Figure 1).

Several problems surfaced during the Stage I process that required a significant amount of time to correct in order to continue production of the Stage I seamless dataset. These include the necessity of coincident stream lines and seed points, overcoming flow direction errors in coastal regions and flat areas, sufficient buffer distances to generate correct flow accumulation and flow direction,

controlling flow direction into and not out of closed basins, and storing several terabytes of data. These situations occurred individually and simultaneously throughout several Cataloging Units.

Stream lines and seed points:

Coincident streamlines and seed points were critical to the Stage I process. If one Cataloging Unit's [stream lines were not coincident with adjacent Cataloging Unit's stream](#) line at the intersection where seed points were placed, seed points were moved upstream or downstream to a place where coincident stream lines existed (Figure 15). The movement of seeds was time consuming because it not only involved the Cataloging Unit that was being processed, but also all Cataloging Units surrounding the one that was being processed. One correction affected numerous areas, and thus the reprocessing of all affected Cataloging Units was necessary. The disadvantage of moving seed points was the deviation from the original digitized 1:250,000 Hydrologic Unit polygon, the huc_poly. However, this was necessary to create a Stage I seamless streamline coverage, and thus seamless coverages for the entire project.

Coastal regions, flat areas, and buffers:

At times, coincident streamlines did not exist at all as was the case in flat and coastal areas. The flow direction per pixel did not have a well-defined path, therefore, Stage I stream lines resulted in [perpendicular stream lines in one](#)

[Cataloging Unit versus an adjacent Cataloging Unit](#) (Figure 16). In these situations, movement of seed points was also necessary to maintain the seamless streamline coverage, and again, the deviation from the 1:250,000 Hydrologic Unit polygon. Sometimes this problem was overcome by rerunning the semi-blind pass process at a larger buffer, such as a 50,000 or 80,000 meter buffer. The larger buffer allowed for relief to correct the flow accumulation so the flow direction would be corrected.

Closed basins:

Closed basins are Cataloging Units where no flow drains out of the unit, however, flow may or may not come into the unit. In an open basin Cataloging Unit, the filling process allows flow only toward the edge of the DEM. In the case of a closed basin, [locate points](#) were instrumental in preventing a “true sink” within the Cataloging Unit from “filling up” and spilling off the edge of the DEM (Figure 14). These points were placed wherever a “true sink” existed within the Cataloging Unit to allow flow to occur not only toward the edge of the DEM, but also through the locate points.

Stage I processing of closed basins required intensive research to reach a determination whether a Cataloging Unit should be processed as an open or closed basin. The “smartlinks” coverage, showing connectivity from the centroid of one Cataloging Unit to the centroid of another Cataloging Unit, was used to determine

whether or not a Cataloging Unit was, in fact, closed. The United States Geological Survey (USGS) Water-Supply Paper 2294, [Hydrologic Units Maps](#) (Seaber, et-al, 1994) was another valuable resource in the determination of closed basins. [NHD](#) coverages were utilized to visually determine if synthetic flow stopped within a Cataloging Unit.

The processing of closed basins was meticulous. Prior to processing a Cataloging Unit, it was imperative to determine if it was truly a closed basin. If a Cataloging Unit was processed as a closed basin when, in fact, it was an open basin, then incorrect placement of locate points within the watershed would “punch” holes through the DEM, and as a result data would be lost.

Managing, storing, and distributing the data

The [EDNA](#) project requires adequate storage space for the terabytes of data produced. The average size of a Cataloging Unit is 150 megabytes. With over 2000 Cataloging Units, excluding Alaska, the total storage space required for all stages of EDNA is at a minimum of 1.3 terabytes. These data are stored and managed on a Sun Ultra-Enterprise server on five separate disks, in addition to, archive storage on a silo tower, tapes and CDs from EDNA Stage I cooperators, and incremental backup storage of the processed data on the Sun server.

The [EDNA Stage I seamless dataset](#) will be loaded, stored, and managed in [ArcSDE](#). This Relational Database Management System (RDMS) utilizes Oracle

8.1.7, and will provide an environment for multi-users to view the data via clients such as [ArcExplorer](#), [ArcView](#), [ArcMap](#), or [ArcIMS](#). The building of pyramids, the creation of statistics, and versioning will allow the end user to access, view, analyze, and download portions of the dataset.

The loading processes for vector coverages were fully automated using AML. Procedures resulting from a Cooperative Research and Development Agreement (CRADA) with [Environmental Systems Research Institute, Inc.](#) (ESRI) were used to load and mosaic the [EDNA raster layers into ArcSDE](#).

Conclusion

Development of the EDNA dataset required intense planning, research, production time, and storage space to produce a national seamless hydrologically derived watershed dataset. Stage I production of EDNA has been implemented by the [USGS](#) through application of Geographic Information Systems, and the assistance of a consortium of participants. Considering the magnitude and scope of [EDNA](#), the production and management of the Stage I data have been successful. Utilization of ArcSDE will greatly enhance the availability of the data layers that contain valuable topological information that will lend itself to many regional to national applications.

References

- Jenson, S.K. and J.O. Domingue, 1988. Extracting topographic structure from digital elevation data for geographic information system analysis: *Photogrammetric Engineering and Remote Sensing*, Vol. 54, pp. 1593-1600.
- Kelly, G., 2000. Description of Stage I NED-H processing and documentation of the output from Stage I Ned-H Processing, 1.
- Seaber, Paul R., F. Paul Kapinos, and George L. Knapp, 1994. Hydrologic Unit Maps. United States Geological Survey Water-Supply Paper 2294. U.S. GPO: Washington, D.C.
- Verdin, K.L., and J.P. Verdin, 1999. A topological system for delineation and codification of the Earth's river basins: *Journal of Hydrology*, Vol. 218, pp.1-12.



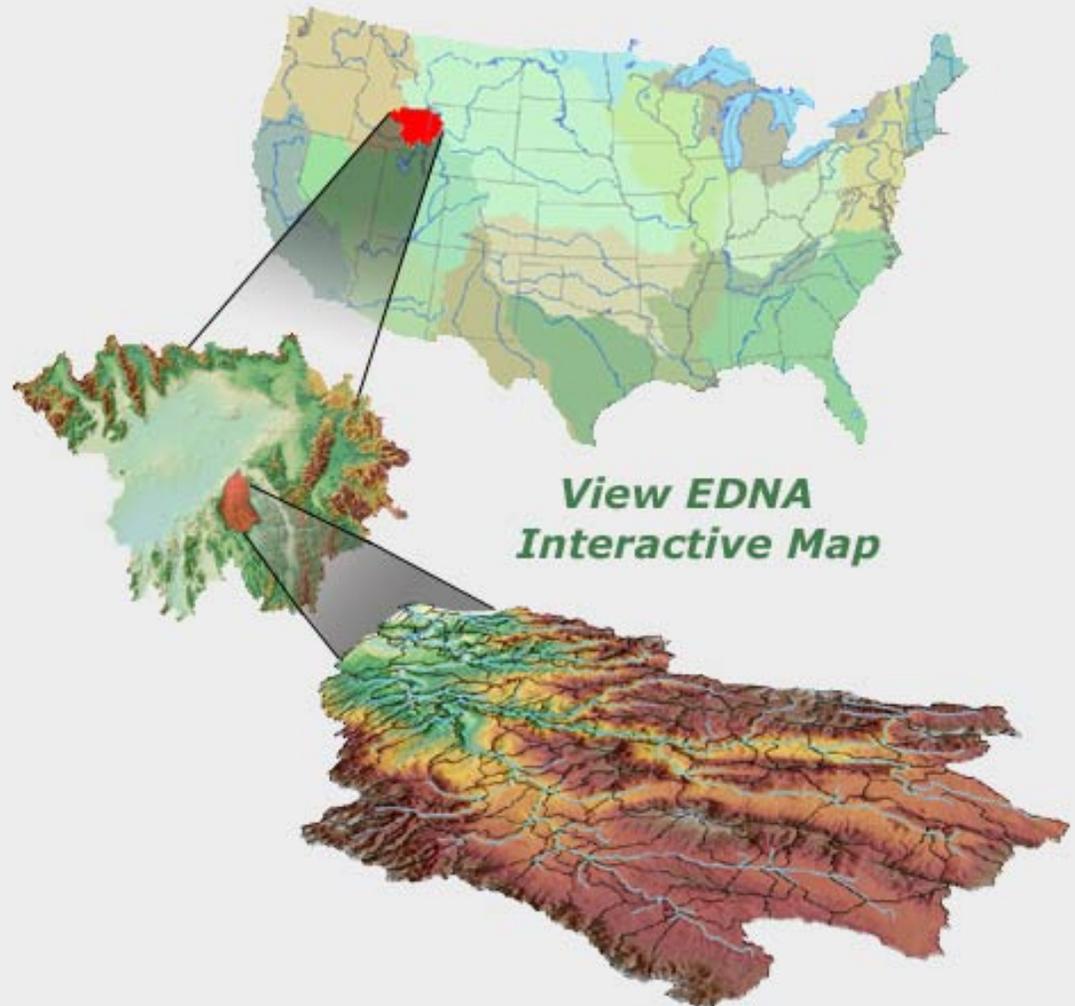
EROS Data Center

Elevation Derivatives for National Applications (EDNA)

EDNA is a multi-layered database derived from a version of the [National Elevation Dataset \(NED\)](#), which has been hydrologically conditioned for improved hydrologic flow representation. The seamless EDNA database provides 30 meters resolution raster and vector data layers including:

- Aspect
- Contours
- Filled DEM
- Flow Accumulation
- Flow Direction
- Reach Catchment Seedpoints
- Reach Catchments
- Shaded Relief
- Sinks
- Slope
- Synthetic Streamlines

Hydrologically conditioned elevation data, systematically and consistently processed to create hydrologic derivatives, can be useful in many topologically based visualization and investigative applications. Drainage areas upstream or downstream from any location can be accurately traced facilitating flood analysis investigations, pollution studies, and hydroelectric power generation projects.



Elevation Derivatives for National Applications

Please read this general [Disclaimer](#)

[U.S. Department of the Interior](#)

[U. S. Geological Survey](#)

National Mapping Division

[EROS Data Center](#)

[Comments and Suggestions](#)

Last modified: Monday, August 6, 2001

<http://gisdata.usgs.gov/edna/index.asp>



You have reached the old NED home Page. You should automatically be forwarded to [default.asp](#) (the new home page).

