Open-File Report 2004-1099
Building a Geospatial ROMA Project Database

by
Robert G. Clark, Steven Kambly, Thomas Moreland, and Milan Pavich

U.S. Department of Interior
U.S. Geological Survey

Building a Geospatial ROMA Project Database

by
Robert G. Clark, Steven Kambly, Thomas Moreland,
and Milan Pavich

Introduction

This Open-File report describes the accomplishments and ongoing activities of the River Observatories for Management Applications (ROMA) Project within the Geography Discipline. This report gives an overview of data-related activities including acquisition, processing, and analysis. The first in a series of three articles scheduled to be published by the Geography Discipline, it is organized as follows: background, areas of study, data descriptions, data sources, systems configuration, geospatial data processing, building datasets for analysis, future data analysis, and conclusions. This report is intended for USGS scientists with knowledge of the USGS BASIS+ System, spatial analysis, and the World Wide Web.

Background

In 2002, the Geography Discipline was asked to participate in land cover and land use studies as a task under the ROMA Project led by Milan Pavich. In 2003, this project had several funding sources: the Chesapeake Bay Program (Place Based Studies), the Earth Surface Dynamics Program, and the Geographic Analysis and Monitoring Program. The USGS corporate project tracking system BASIS+ discusses these and other program issues. In BASIS+, the ROMA Project objectives are defined as follows:

> "The overarching ROMA Project is to bring together USGS, university, and other agency researchers to conduct studies of the impacts of sediment fluxes on habitats in various parts of the Susquehanna River Basin. The significance of sediment to habitat restoration and water quality has recently been highlighted in creation of the Chesapeake Bay program. The Susquehanna River is historically the largest single sediment source for Chesapeake Bay. The objective in this project is to identify the main sources of sediment to the Susquehanna by quantifying the sediment yields from tributary basins and the contributions of disturbed soil surfaces to the sediment budgets."

The following paragraph is a description of the Geography Discipline Task entitled GIS Analysis of Consequences of Land Cover/ Land Use Change:

> "The uplands of the Susquehanna watershed have been changed by human-induced and natural processes, some of which have significant impact on ecosystem health and sustainability. The need to integrate and apply information to help understand the consequences of land surface change on sediment erosion and deposition is critical to managing the natural resources of the watershed and Chesapeake Bay. Land surface change may result from agricultural production, urbanization, forest logging, climate change, and other factors operating at local and broad regional scales. Improved information and understanding about the state of the land surface and the rates and patterns, causes/drivers, and

consequences of landscape change are needed to help scientists and decision-makers with land-use planning, land management, and natural resource utilization/conservation. For example, accurate studies of sediment mobilization (a critical area of investigation to the Bay program) require land cover data and high resolution elevation data linked to detailed slope, aspect, and accumulated flow calculations in order to model the sediment carrying ability of local sub-watersheds. Current studies (Langland and others, 1999) have quantified sediment accumulation behind four dams on the lower Susquehanna using less than satisfactory base cartographic data. Current available elevation data are not consistent or detailed enough in resolution in the area of needed coverage to factor in local topography and land cover effects to satisfactorily predict sediment movement."

Areas of Study

Part of the project's formulation defined four nested study areas of varying scales. The study areas, shown with example datasets in **Figure 1,** lie within the Chesapeake Bay Watershed. The Susquehanna Watershed is the largest study area and comprises the northern half of the Chesapeake Bay Watershed. The second study area, Lancaster County, PA, falls mostly within the Susquehanna Watershed. Little Conestoga Creek Watershed which lies within the boundaries of Lancaster County is the third study area. The West Branch Watershed of the Little Conestoga Creek Watershed is the final study area and the smallest. A wide variety of data were collected for these four study areas.

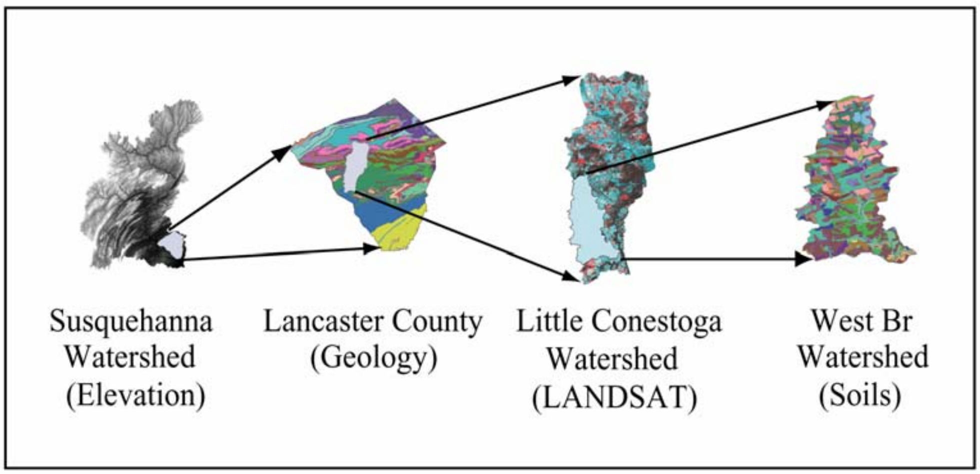


Figure 1. Four Study Areas with Example Datasets

The study areas comprise a highly productive agricultural region that is undergoing urbanization near population centers. Therefore, the analysis of land surface change and its associated impacts on streams and rivers requires many different kinds of data to capture the types and rates of change. These data were collected from several sources and varied in scale and resolution depending on the needs of the original data producers who include county, state, and national agencies. As a result, the project's four hierarchical study areas have inherent data continuity issues, especially when vertically integrated.

Data Descriptions

At the beginning of the project, project personnel decided that common base geospatial data were needed for the four study areas. These data include National Land Cover Dataset (NLCD), Digital Orthophoto Quadrangles (DOQ), Digital Elevation Models (DEM) such as the National Elevation Dataset (NED), Digital Raster Graphics (DRG), and National Hydrography Dataset (NHD). Other important datasets collected include Soil Survey Geographic (SSURGO) Data Base, State Soils Geographic (STATSGO) soils, and Surface Geology. Subsequent ROMA Land Cover Change Analyses will include National High Altitude Photography (NHAP), National Aerial Photography Program (NAPP), additional years of LANDSAT, and Lancaster County Land Use and Impervious Surface datasets.

Table 1 is a snapshot of a data inventory spreadsheet listing the important data attributes and the status of data collection and processing as column titles. The researcher's are continually updating this multi-year project data. This ongoing work and its data will be used in future project analysis.

			Dataset	Inventory			
Susquehanna	**Type of Data**	**Data Structure**	**Format**	**Projection**	**Scale**	**Cell Size**	**Producer**
	HUC	Polygon	ESRI	Albers 83			WRD
	Soil/STATSGO	Polygon	Shapefile	Albers 83	250K		NRCS
	Geology	Polygon	Shapefile	Albers 83	2.5Mil		GD
	Ecoregion	Polygon	Shapefile	Albers 83			NMD
	NLCD	Raster	ESRI/grid	Albers 83		30M	NMD
	Elev (DEM)	Raster	ESRI/grid	Albers 83		30M	NMD/GDA
	Shade Relief	Raster	ESRI/grid	Albers 83		30M	NMD
	Roads	Line	ESRI	Albers 83	100K		NMD/Atlas
	Erosion Index Stations	Point	ESRI	Albers 83			Pavich
Lancaster County	**Type of Data**	**Data Structure**	**Format**	**Projection**	**Scale**	**Cell Size**	**Producer**
	Co Boundary	Polygon	ESRI	Albers 83			Geol Poly
	Geology	Polygon	ESRI	Albers 83	2.5Mil		PASDA
	Soil/STATSGO	Polygon	Shapefile	Albers 83	250K		NRCS
	Soil/SSURGO	Polygon	Shapefile	Albers 83	63K		NRCS
2acquire	NHD - Drain	Polygon	ESRI	Albers 83	24K		NMD
2acquire	SPOT	Raster		Albers 83		20M	EURO
Little Conestoga	**Type of Data**	**Data Structure**	**Format**	**Projection**	**Scale**	**Cell Size**	**Producer**
	HUC	Polygon	ESRI	Albers 83			WRD
	Soil/SSURGO	Polygon	Shapefile	Albers 83	63K		NRCS
	Geology	Polygon	Shapefile	Albers 83			
	NLCD	Raster	ERDAS/img	Albers 83		30M	NMD

	DRG	Raster	ERDAS/img	Albers 83	24K		NMD
	DOQ	Raster	ERDAS/img	Albers 83		1M	NRCS
	Elev (NED)	Raster	ERDAS/img	Albers 83		10M	NMD/GDA
	Slope	Raster	ERDAS/img	Albers 83		10M	Generated
	Shade Relief	Raster	ERDAS/img	Albers 83		10M	NMD
	LANDSAT 7	Raster	ERDAS/img	Albers 83		30M	NMD/EDC
	NHD - Drain	Polygon	ESRI	Albers 83	24K		NMD
In work	Co. Land use			**redo**			PASDA
2do	Aspect	Raster	ERDAS/img	Albers 83		10M	Generated
2do	Regrid NED	Raster	ERDAS/img	Albers 83		Appr.2.5M	NMD
2acquire	LIDAR	Raster	ERDAS/img	Albers 83		1M+	
2acquire	SPOT	Raster				20M	EURO
West Branch	**Type of Data**	**Data Structure**	**Format**	**Projection**	**Scale**	**Cell Size**	**Producer**
	HUC	Polygon	ESRI	Albers 83			WRD
	Soil/SSURGO	Polygon	Shapefile	Albers 83	63K		NRCS
	NLCD	Raster	ERDAS/img	Albers 83		30M	NMD
	Shade Relief	Raster	ERDAS/img	Albers 83		10M	NMD
	Elev (NED)	Raster	ERDAS/img	Albers 83		10M	NMD/GDA
	Slope	Raster	ERDAS/img	Albers 83		10M	Generated
	LANDSAT 7	Raster	ERDAS/img	Albers 83		30M	NMD/EDC
	NHD - Drain	Polygon	ESRI	Albers 83	24K		NMD
In work	Co. Land use			**redo**			PASDA
In work	DOQ	Raster	ERDAS/img	Albers 83		1M	NRCS

Table 1. Sample Data Inventory Spreadsheet

Data Sources

Project participants collected many kinds of geospatial data from several data servers: *The National Map* (USGS), the Geospatial Data Architecture (USGS), the Department of Agriculture's Natural Resources Conservation Service (NRCS) Geospatial Data Gateway, and Pennsylvania State University's Pennsylvania Spatial Data Access (PASDA).

"*The National Map* is a consistent framework for geographic knowledge needed by the Nation. It provides public access to high-quality, geospatial data and information from multiple partners to help support decision-making by resource managers and the public. *The National Map* is the product of a consortium of Federal, State, and local partners who provide geospatial data to enhance America's ability to access, integrate, and apply geospatial data at global, national, and local scales. The U.S. Geological Survey (USGS) is committed to meeting the Nation's needs for current base geographic data and maps. Our vision is that, by working with partners, we will ensure that the Nation has access to current, accurate, and nationally consistent digital data and topographic maps derived from those data." (see below) The Figure 2 is a screen capture of *The National Map*'s Web Homepage. It can be accessed at http://nationalmap.usgs.gov/.

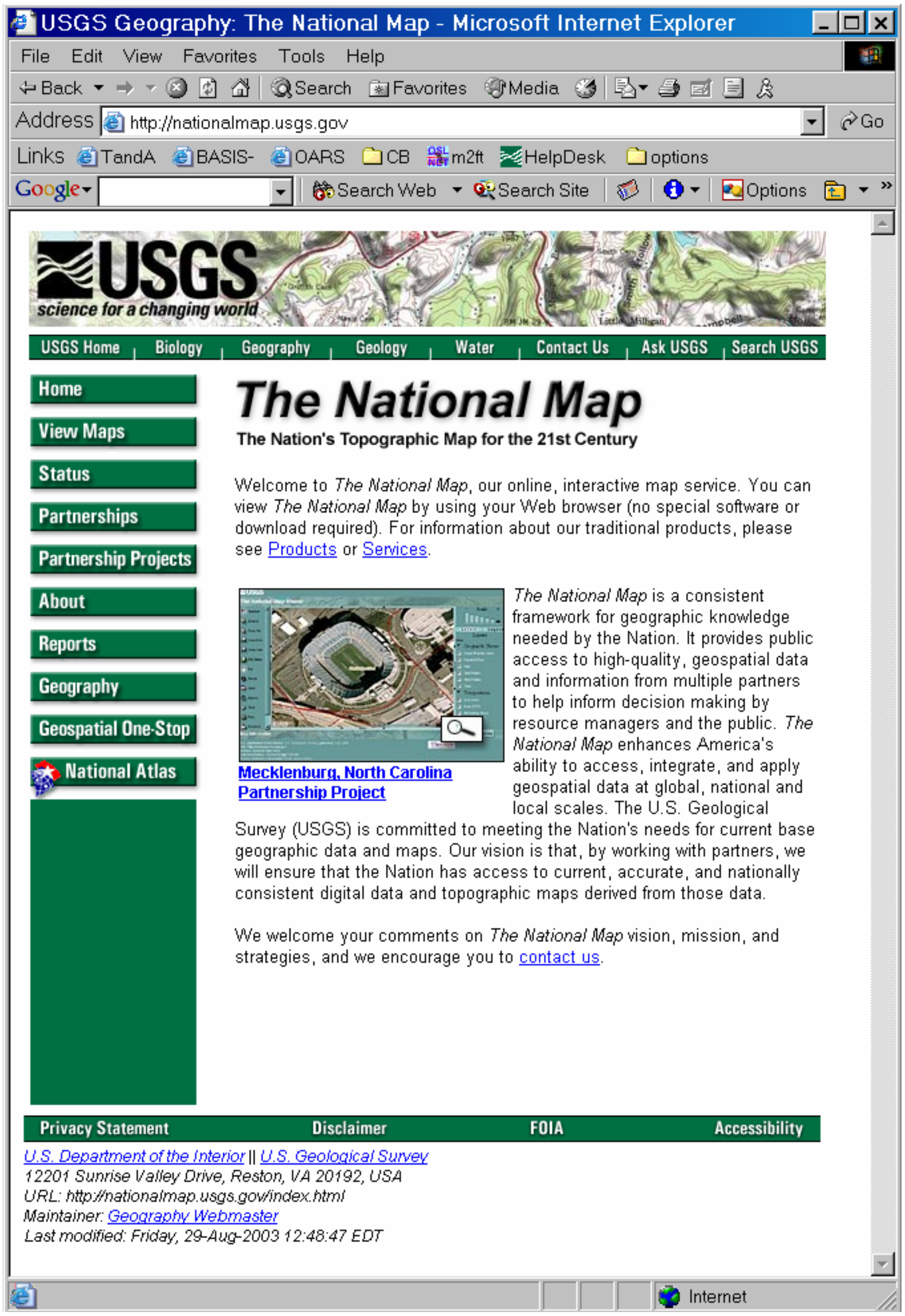


Figure 2. *The National Map's* Homepage

The GDA is a database application developed by the Geography Discipline that retrieves archived USGS geospatial data and metadata. Users may query and download its data to a client workstation. The GDA server is located at EROS Data Center so data transfer rates are often very good. **Figure 3** is a screen capture of how the GDA client looks on a typical workstation.

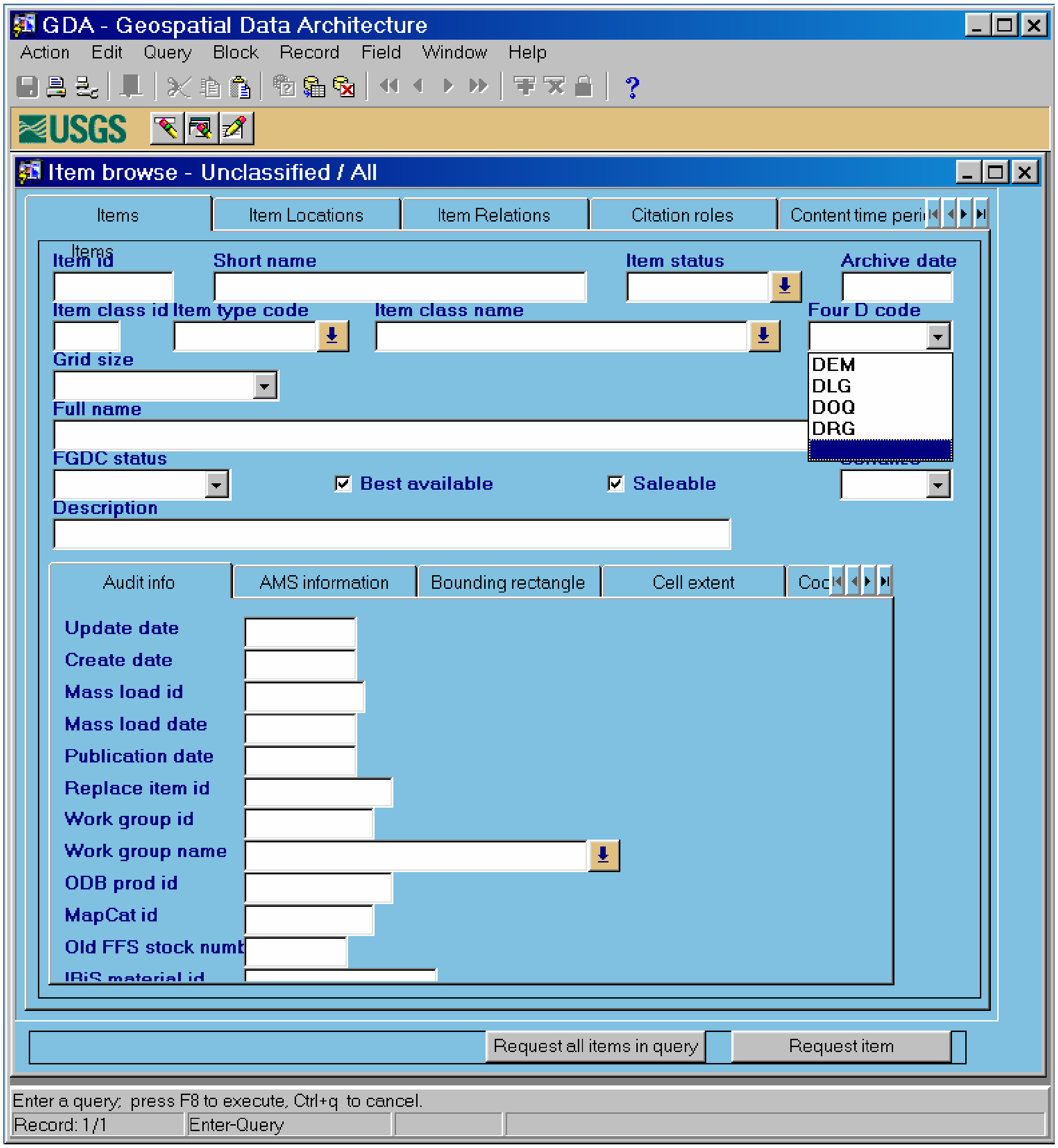


Figure 3. GDA Client Screen

Another source of data was the National Resource Conservation Service through their Geospatial Data Gateway application (see http://lighthouse.nrcs.usda.gov/gateway/NextPage.asp). The Geospatial Data Gateway has a very user-friendly client interface which guides the user through data query and download in a step-by-step fashion as is shown in **Figure 4.**

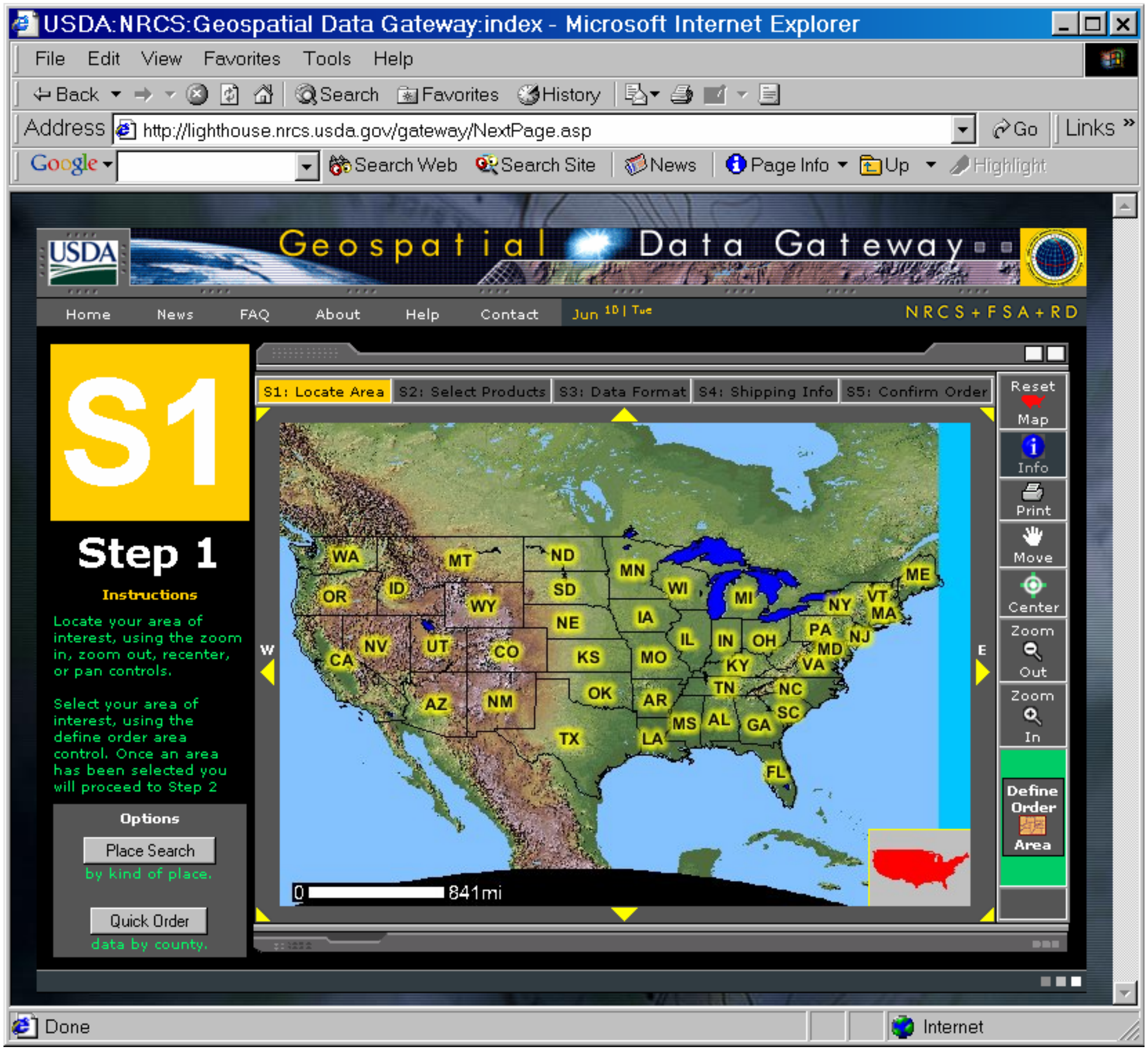
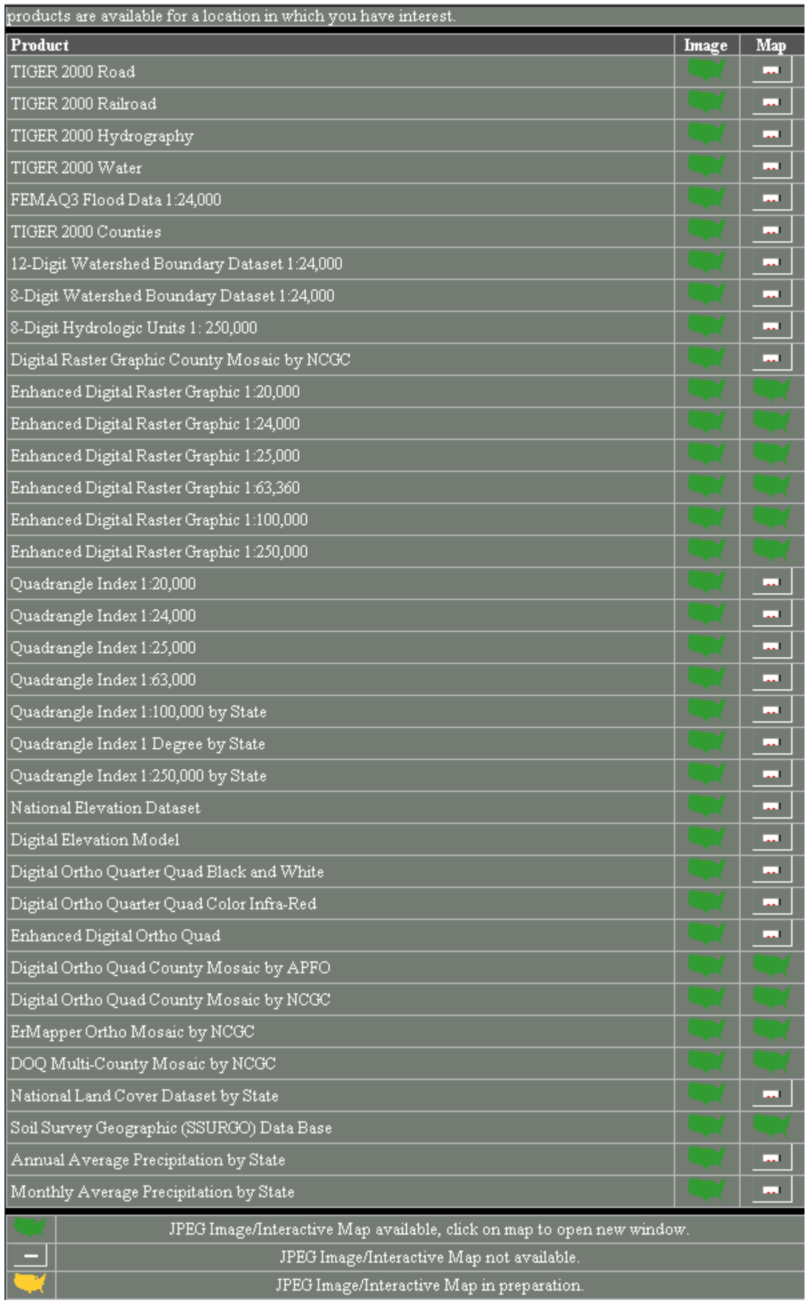


Figure 4. Geospatial Data Gateway Screen

In addition, figure 5 illustrates the breadth of currently available data at this. (see http://lighthouse.nrcs.usda.gov/gateway/statusmaps.asp).



products are available for a location in which you have interest.

Product	Image	Map
TIGER 2000 Road		
TIGER 2000 Railroad		
TIGER 2000 Hydrography		
TIGER 2000 Water		
FEMAQ3 Flood Data 1:24,000		
TIGER 2000 Counties		
12-Digit Watershed Boundary Dataset 1:24,000		
8-Digit Watershed Boundary Dataset 1:24,000		
8-Digit Hydrologic Units 1: 250,000		
Digital Raster Graphic County Mosaic by NCGC		
Enhanced Digital Raster Graphic 1:20,000		
Enhanced Digital Raster Graphic 1:24,000		
Enhanced Digital Raster Graphic 1:25,000		
Enhanced Digital Raster Graphic 1:63,360		
Enhanced Digital Raster Graphic 1:100,000		
Enhanced Digital Raster Graphic 1:250,000		
Quadrangle Index 1:20,000		
Quadrangle Index 1:24,000		
Quadrangle Index 1:25,000		
Quadrangle Index 1:63,000		
Quadrangle Index 1:100,000 by State		
Quadrangle Index 1 Degree by State		
Quadrangle Index 1:250,000 by State		
National Elevation Dataset		
Digital Elevation Model		
Digital Ortho Quarter Quad Black and White		
Digital Ortho Quarter Quad Color Infra-Red		
Enhanced Digital Ortho Quad		
Digital Ortho Quad County Mosaic by APFO		
Digital Ortho Quad County Mosaic by NCGC		
ErMapper Ortho Mosaic by NCGC		
DOQ Multi-County Mosaic by NCGC		
National Land Cover Dataset by State		
Soil Survey Geographic (SSURGO) Data Base		
Annual Average Precipitation by State		
Monthly Average Precipitation by State		

JPEG Image/Interactive Map available, click on map to open new window.

JPEG Image/Interactive Map not available.

JPEG Image/Interactive Map in preparation.

Figure 5. Geospatial Available Data

Once a user selects Geospatial Available Data he/she receives an email specifying where and when data can be acquired by a browser/ftp process, including the exact browser command to download the data. Another browser enabled data server is Pennsylvania State University's PASDA, the Pennsylvania Geospatial Data Clearinghouse (see http://www.pasda.psu.edu/). The data available from this site are primarily of the State of Pennsylvania and surrounding areas. This data server also has links to other important data sources such as the Chesapeake Bay Program Data Catalog. **Figure 6** illustrates some of the data available from this server.

Figure 6. Accessing PASDA Data

These four data severs were important to the free or "public domain" data collection efforts of the ROMA project; however, *The National Map* (see http://nationalmap.usgs.gov/) is becoming the data server of choice by the US Geological Survey research community as a source of the highest quality geospatial data for project spatial data processing.

Systems Configuration

The research team considered several critical to develop the specifications for the project's computer systems including hardware, operating system, network software and applications software.

The project uses four desktop workstations configured with half a gigabyte (GB) of memory and 1.8 gigahertz central processing units. All systems have a 33 GB disk drive; however one system has an additional 120 GB drive that it shares over the network with the other workstations. In addition, each system has a graphics card with at least 32 megabytes of memory and a twenty-one inch monitor. This system configuration is adequate to manipulate and process the project's large datasets.

The Windows operating system was installed on the project workstations. The researchers configured the network connection, My Network Places, to run over the USGS National Center's 100 Megabit Ethernet. The large 120 GB disk drive was shared using the My Network Places protocol and appeared as a local drive on the other systems. Using this simple regional approach, all Eastern Region Geography Project workstations shared a common geospatial database without incurring excessive administrative costs.

The application software was also distributed through the network. Workstation network licenses for both Geographic Information System (GIS) software and Image Processing software were used, saving software funds. Project personnel used several GIS and image processing systems to manipulate, integrate, and analyze both vector and raster datasets.

With the system and network configurations described above, scientists were able to efficiently acquire, manipulate and analyze geospatial data for the ROMA Project.

Geospatial Data Processing

Once the data were collected, they were processed and prepared for spatial analysis. Data preparation involved several steps—mosaicking, clipping, re-projecting, and format converting. All of these processes were performed using commercially available software. Raster data were processed using an image processing system; vector data were processed with a GIS.

The mosaic operation involved selecting adjacent images and cropping borders where appropriate. Then task scientists stitched the images together using the appropriate algorithms. Depending on the size of the images, the mosaic operation could involve relatively long processing times (ERDAS 2001).

After a mosaic was assembled, typically the scientist performed a clip or subset procedure. Clipping, "cookie cutter" operation used to extract a desired study area from a larger dataset, is an automated procedure available in many GIS and image processing applications. Generally, the user starts by specifying two input files: a file having the desired shape or boundary and a file containing the desired type of data for that bound area. The user then specifies the name of the output layer and then executes the program to perform the clipping operation.

Three of the study areas were clipped by Hydrological Unit Code (HUC) boundaries. A HUC is an irregular polygon based on characteristics of a particular watershed boundary (Seaber 1987), (see http://water.usgs.gov/GIS/huc.html); however, task scientists clipped the Lancaster County study area was clipped using county boundaries. In **Figure 7**, the West Branch HUC was clipped from the Little Conestoga Creek HUC using a common GIS software package. **Figure 1** illustrates the spatial extents of all four study areas.

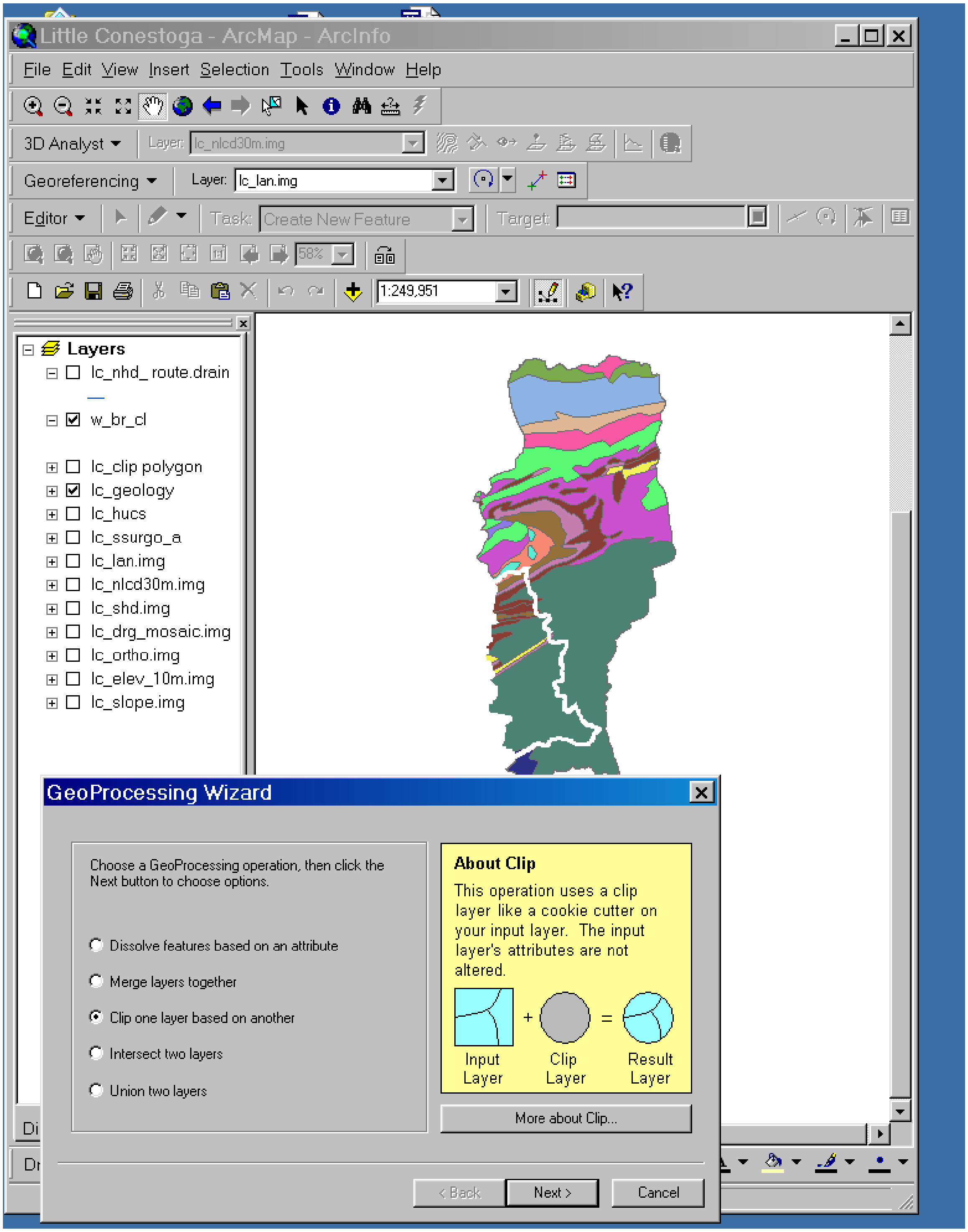


Figure 7. West Branch HUC

Scientist extracts both raster and vector datasets using the cookie-cutter process outlined above. Other methods of data clipping are available.

After the clipping operation was performed, scientists re-projected the data to the Albers Conic Equal Area projection. Researchers choose Albers Conic Equal Area to ensure compatibility with the National Land Cover Trends Project whose methods for land cover interpretation will be used in the ROMA project (see http://landcover.usgs.gov/landcovertrends.html). As in all the data processing operations, raster data were projected in an image processing system while vector data were re-projected using a GIS.

Both GIS and image processing systems have fairly extensive file format conversion routines. Task scientists frequently used several of the routines shown below in **Figure 8** for file conversions**.**

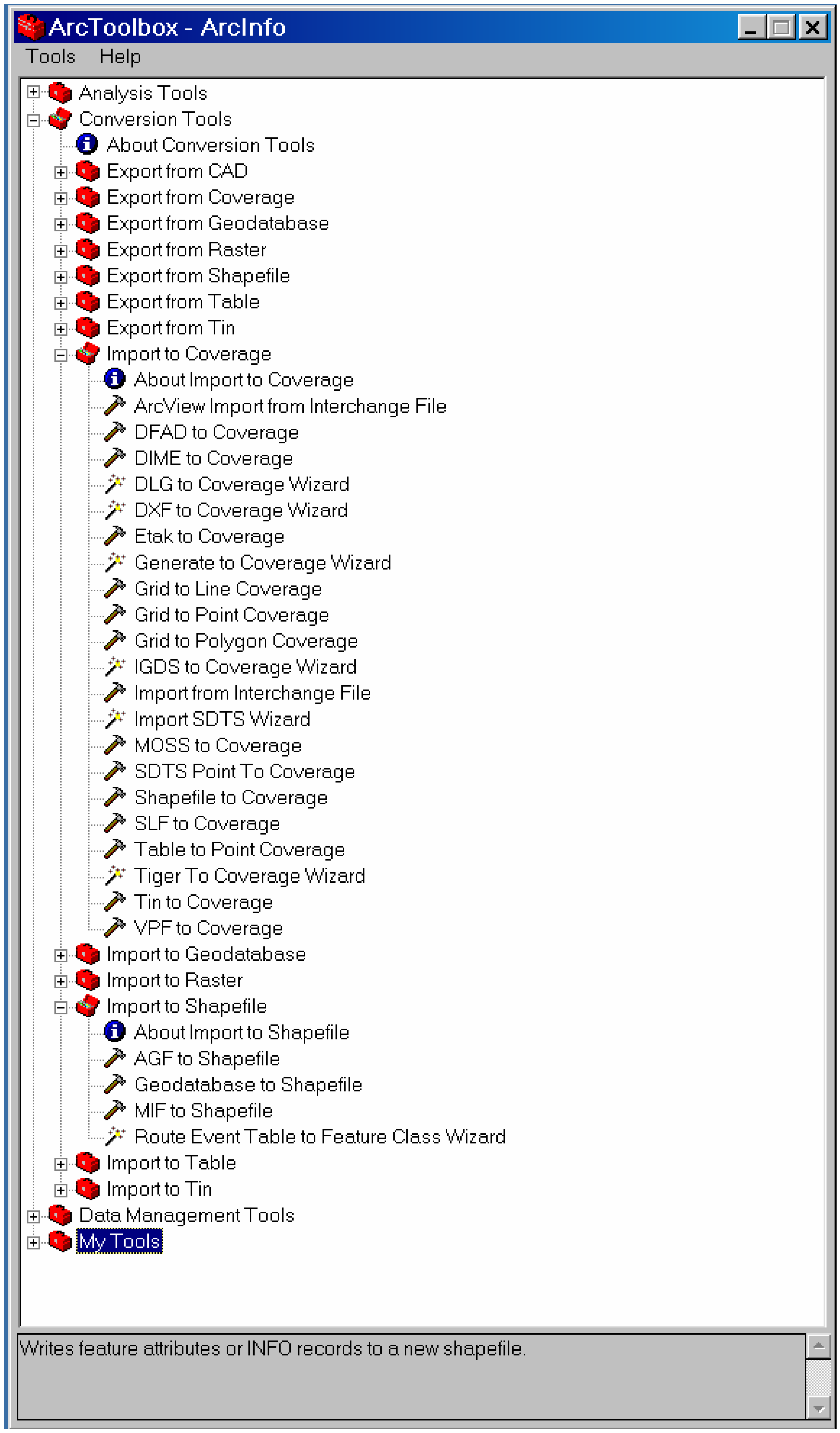


Figure 8. File Format Conversion Routines

There were many more spatial operators employed by Project scientists. The above are only a few examples. Other spatial routines like buffering and Boolean operators are currently being used in data analysis.

Building Datasets for Analysis

Subsequent Open-File reports will describe additional datasets used in analysis of land cover and land use change, and analysis and modeling of sediment mobilization in the Little Conestoga Creek. In addition, project scientist will perform analysis of soil erosion and sediment tracing using beryllium-10 and correlate with sediment movement. The use of beryllium-10 has been shown effective in dating stream sediments (Brown et al, 1988; Valette-Silver et al., 1986).

Task scientists have performed preliminary analysis using derived datasets such as a slope dataset. For example, a slope dataset was derived from elevation data by using an image processing system routine. Next, a soils dataset was set to fifty percent transparency and overlaid on the slope dataset as illustrated in **Figure 9**. This merged dataset will be used to illustrate where certain slopes for a particular soil type are most susceptible to erosion. Scientists will continue to analysis this image and report in later publications.

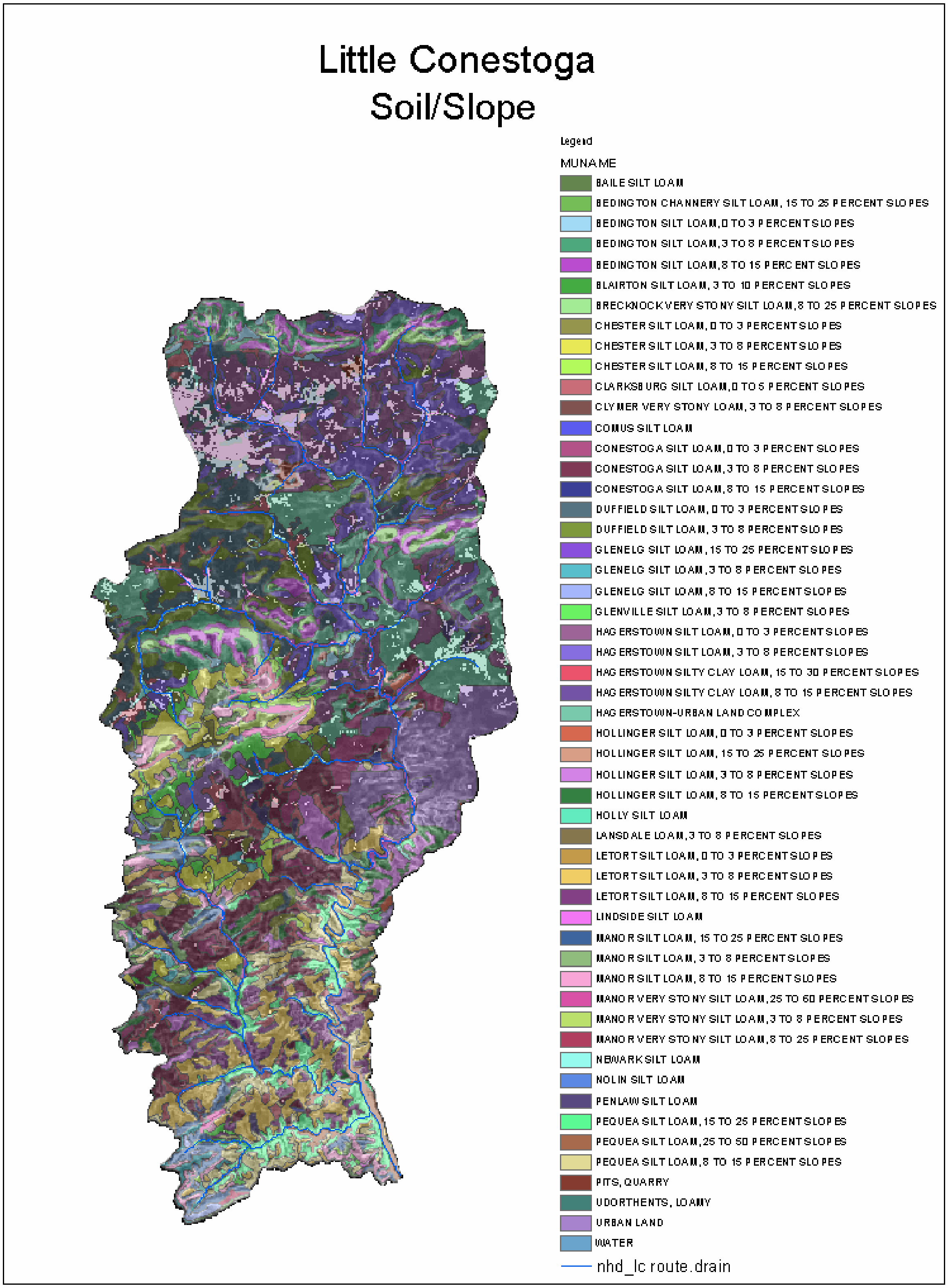


Figure 9. Soil/Slope Overlay Dataset

Future Data Analysis

Additional spatial analysis, using these datasets, will include land cover and land use change analysis and modeling of stream sediment movement as a consequence of land cover and land use change. Task scientists will analyze data such as Digital Raster Graphics (DRGs), Digital Orthophoto Quads (DOQs), scanned NAPP and NHAP aerial photography, NLCD 2001, a new dataset, and multiple years of LANDSAT to determine the land cover change that has occurred from 1969 to 2002. Modeling of sediment mobilization for a specific basin will use the image processing systems, GIS, and the Water Erosion Prediction Project (WEPP) modeling software. Scientists will use soils, slope, and impervious surface data in modeling and analysis. The results of both analyses will be documented in professional papers and/or Open-File Reports.

Conclusions

Data collection and processing are the essential first steps for most research projects. This paper summarizes these activities for the ROMA Project. Optimally, the project datasets should be distributed on a website such as the Chesapeake Bay Web server. See (http://www.chesapeakebay.net/). The data collection, manipulation and preliminary analysis activities detailed here provide a foundation for additional geographic analysis and contributions to science.

References

Environmental Systems Research Institute, Inc., 2000. Using ArcToolBox: Redlands, Calif., Environmental Systems Research Institute, Inc., 30 p.

Brown, L., Pavich, M.J., Hickman, R.E., Klein, J. and Middleton, R., 1988. Erosion of the eastern United States observed with Be-10, Earth Surface Processes and Landforms, 13: 441-457.

ERDAS, Inc., 2001. ERDAS IMAGINE Tour Guide: Atlanta, GA, 304 p.

Langland, M.J., Blomquist, J.D., Sprague, L.A., Edwards, R.E., 1999. Trends and Status of Flow, Nutrients, and Sediments for Selected Nontidal Sites in the Chesapeake Bay Watershed, 1985-98: U.S. Geological Survey Open File Report 99-451, variously paged

Loveland, T.R., Sohl, T.L., Stehman, S.V., Gallant, A.L., Sayler, K.L., and Napton, D.E., 2002. A Strategy for Estimating the Rates of Recent United States Land-cover Changes, *Photogrammetric Engineering and Remotes Sensing*, 68(10): 1091-1100.

Seaber, P.R., Kapinos, F.P., and Knapp, G.L., 1987. Hydrologic Unit Maps: U.S. Geological Survey Water-Supply Paper 2294, 63 p.

Valette-Silver, J.N., Brown, L., Pavich, M.J., Klein, J. and Middleton, R., 1986. Detection of erosion events using Be-10 profiles; examples of the impact of agriculture on soil erosion in the Chesapeake Bay area (USA), Earth and Planetary Science letters, 80, 82-90.