National Hydrography Dataset—Linear Referencing

Linking Information About Geospatial Data

Geospatial data normally have a certain set of standard attributes, such as an identification number, the type of feature, and name of the feature. These standard attributes are typically embedded into the default attribute table, which is directly linked to the geospatial features. However, it is impractical to embed too much information because it can create a complex, inflexible, and hard to maintain geospatial dataset.

Many scientists prefer to create a modular, or relational, data design where the information about the features is stored and maintained separately, then *linked* to the geospatial data. For example, information about the water chemistry of a lake can be maintained in a separate file and linked to the lake. A Geographic Information System (GIS) can then relate the water chemistry to the lake and analyze it as one piece of information. For example, the GIS can select all lakes more than 50 acres, with turbidity greater than 1.5 milligrams per liter (mg/L) (fig. 1).

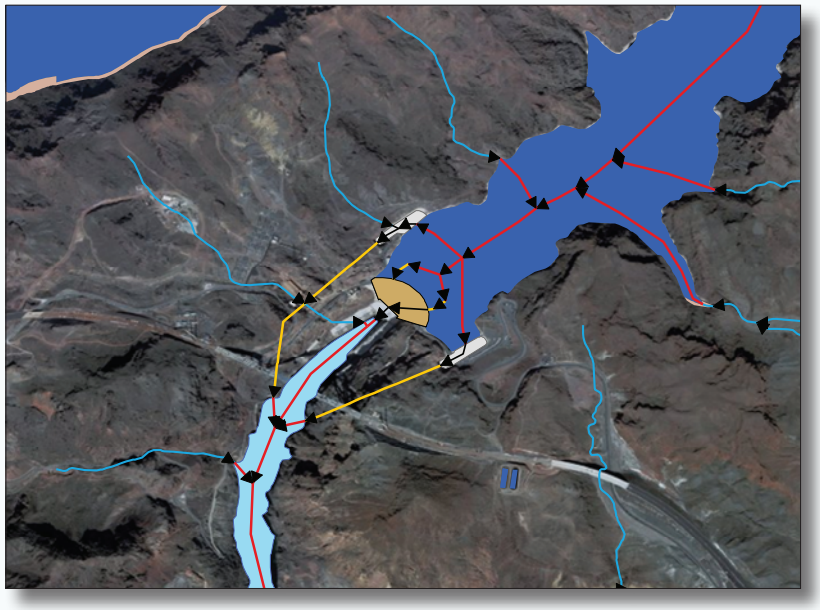

Figure 2. Lake Mead and the Hoover Dam as represented in the National Hydrography Dataset. Flow direction arrows connect this complex network of features, including rivers (blue lines), artificial flowpaths (red lines), and pipelines (yellow lines). Imagery courtesy of USGS 2011.

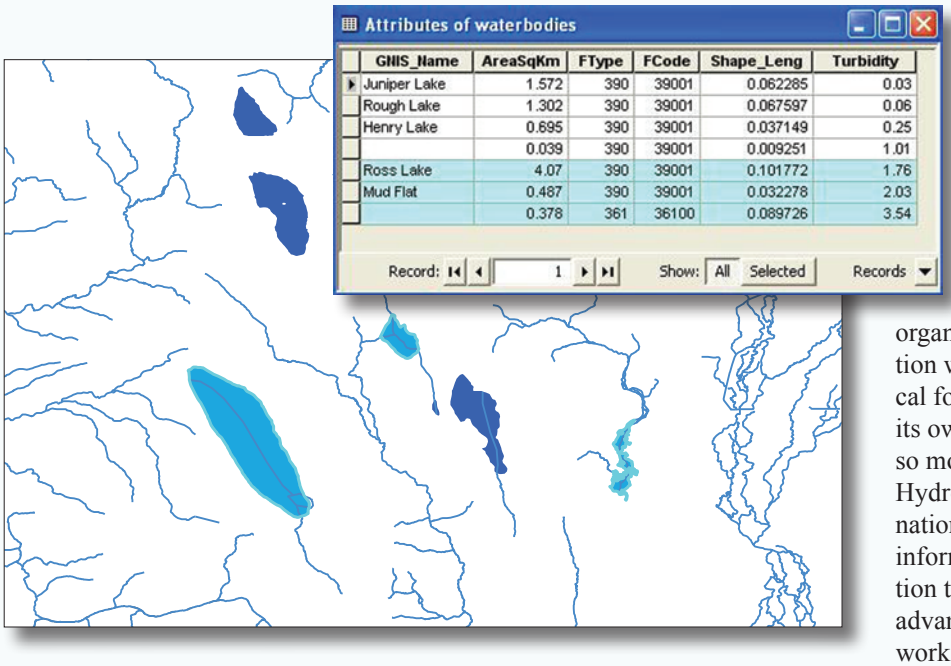

Figure 1. View of map frame showing selected lakes of 50 acres or more in size and with turbidity greater than 1.5 milligrams per liter.

The National Hydrography Dataset

In the analysis of surface-water systems, there is a vast amount of location information that can be linked, such as flow-volume, velocity, temperature, water chemistry, pollution control classifications, aquatic species habitat, recreation designations, or water rights. Such information is typically maintained by various organizations, yet often needs to be used in conjuction with comprehensive analysis. It is impractical for each organization to develop and maintain its own geospatial dataset for locating the data, so most water science agencies use the National Hydrography Dataset (NHD) as the common national framework for surface-water geospatial information (fig. 2). The NHD allows for information to be shared and integrated. The NHD takes advantage of a comprehensive national flow network, provides maintenance through stewardship, and reduces costs by eliminating data redundancy and optimizing systems and tools.

Linear Referencing

Many agencies at Federal, State, and local levels, as well as research, conservation, academic, and industry organizations, produce and analyze water information. Because this information should be shared and integrated, these organizations need a universal system for linking their data. This information can be quite complex, often overlapping multiple features, or representing a small portion of a single feature. And one agency's fish habitat information may not start and end exactly where another agency's water temperature information starts and ends. To allow for these types of information to be linked to the NHD in a practical manner, a system of *linear referencing* is used.

Linear referencing can be used as an addressing system similar to the addressing used to identify the position of a house on a street. Linear referencing is used in NHD to assign reaches and measures to surface-water locations. Where a zip code and street address are used to locate a residence, a ReachCode and Measure are used to locate a surface-water event in NHD.

ReachCodes

To create the "streets," line segments of streams and rivers are labeled with a *ReachCode*. This code is used instead of stream name because most streams are not named or do not have unique names, and streams can be too complex and dynamic for name designators. ReachCodes can account for this naming problem. The ReachCode is a 14-digit number of which the first 8 digits ("zip code") contain the fourth-level hydrologic unit code or zip code where the stream is located. The final six digits are a unique number inside the fourth-level hydrologic unit (fig. 3). ReachCodes are unique. ReachCodes also are applied to lake polygons to ensure they are all uniquely identified.

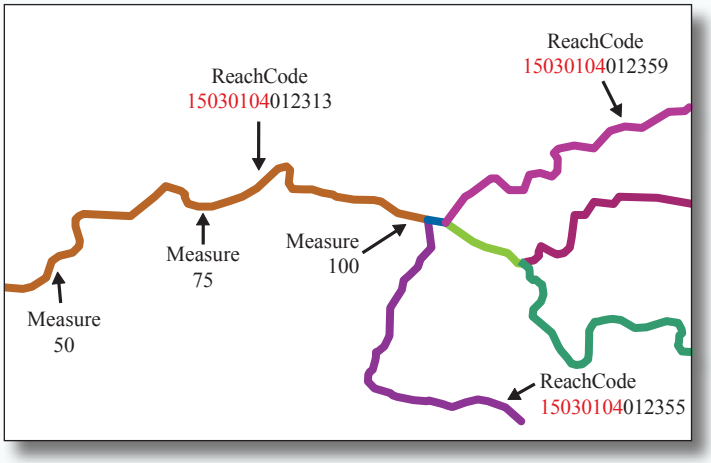

Figure 3. National Hydrography Dataset flowlines symbolized by ReachCodes. The first 8 digits of the ReachCode identify the 8-digit hydrologic unit or subbasin. Measures are expressed as a percent, with the most upstream end of ReachCode with a value of 100 and the most downstream end with a value of 0.

Measure

To locate the position of information on the ReachCode, a value called a *Measure* is used. A Measure is a unitless value ranging from "0" at the downstream end of a ReachCode segment to "100" at the upstream end (fig. 3). No matter how long, short, or sinuous the ReachCode segment, the values always range from 0–100, and up to four decimal places can be used. For example, a streamgage may be located on ReachCode segment 14010002000461 and Measure 19.3815. That means the streamgage is located 19.3815 percent upstream on this particular ReachCode segment. Measures are universal and versatile for use in any complex situation. When locating information on a curvilinear portion of a river, rather than at a point, a *From_Measure* is used to mark the beginning of the information and an *End_Measure* is used to mark the ending location. These Measures can be located on different ReachCodes that are on the same flowpath.

The Power of Location on the Network

As with any geospatial data, it is important to identify the location of surface-water information. Historically, this has often been done with a latitude and longitude, which locates the water information in x-y space on the surface of the earth (fig. 4).

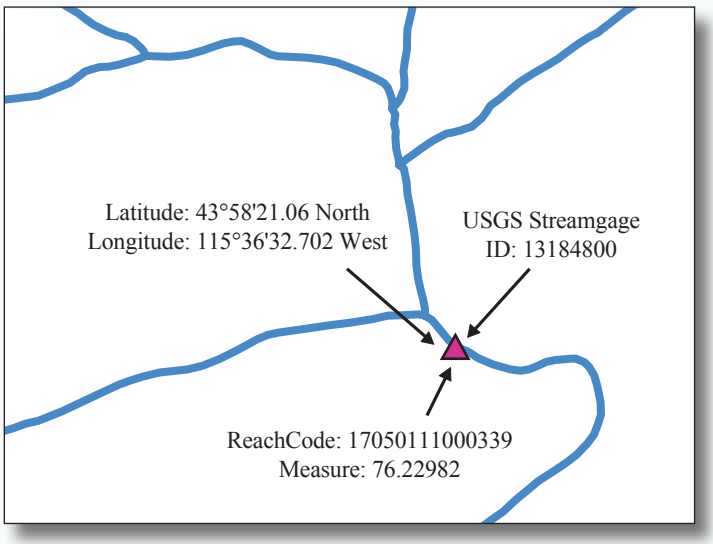

Figure 4. U.S. Geological Survey streamgage 13184800 can be identified with latitude and longitude or on the network at ReachCode 17050111000339 and Measure 76.22982.

However, in the case of surface water, it is much more powerful to locate the water information using linear referencing, as represented by the NHD, because most surface-water information exists within the context of a drainage network, in addition to simple x-y space. The network allows the establishment of upstream and downstream relations of water information. Thus, information, such as streamgages or water diversion locations, can be related to each other on the network, enabling cause and effect analysis. This creates improved data relations that can lead to greater scientific discovery. Using linear referencing to locate water information on the NHD network creates new opportunities for developing an understanding of how the Nation's vast resources of water data can be analyzed in greater depth.

By Jeffrey Simley and Ariel Doumbouya

For more information, contact:
National Hydrography Dataset
U.S. Geological Survey, Box 25046, MS 510, Denver, CO 80225
http://nhd.usgs.gov, email: nhd@usgs.gov