Title
Migrating attributes tied to the EPA river reache file to the national hydrography dataset

Authors
Moore, Cynthia L.
Mullins, James C.
Willett, Karen Beardsley
et al.

Publication Date
1999-02-01

Cynthia L. Moore, James C. Mullins, Karen B. Willett, and James F. Quinn

Migrating Attributes tied to the EPA River Reach file to the National Hydrography Dataset

In cooperation with several federal and state agencies, the Information Center for the Environment (ICE) at the University of California Davis is analyzing the ability to cross-reference information linked to California's hydrography dataset (a modified form of the U.S. Environmental Protection Agency's River Reach File) to the new National Hydrography Dataset. California's Reach File contains a critical but non-standard foreign key that is essential to many research applications. Additionally, ICE plans to test, verify, and recommend enhancements to cross-referencing methods to improve the transfer of reach-code information from earlier hydrography datasets in California to the NHD.

Introduction

Increasingly, both local and national efforts to protect and restore rivers and riparian environments rely upon a common base of spatial data. Organized hydrographic information, especially, is vital to the success of environmental planning and management. Until this year, the only nationwide addressing system for rivers was the U.S. Environmental Protection Agency's (EPA) River Reach File. While the Reach File was developed to support EPA regulatory responsibilities, it has been mandated as the national spatial descriptor for a growing array of topics ranging from water quality to biodiversity. In California, environmental assessment and restoration programs throughout the state have benefited from the standardized Reach File.

The EPA's River Reach File is a geographic and hydrographic database of the surface waters of the continental United States and Hawaii. The structure and content of the Reach File database were created expressly to establish hydrologic ordering, to perform hydrologic navigation for modeling applications, and to provide a unique identifier for each surface water feature (Horn et al 1994). Flow direction and hydrologic ordering are inherent in its connectivity, and this connectivity works regardless of topologic continuity.

The EPA and USGS are committed to further enhancing the River Reach File into a more usable, correct, and stable hydrography base layer. The next step towards this goal, the National Hydrography Dataset (NHD), will provide improved integration of hydrologic data for the United States. NHD, which is based on 1:100,000-scale, Digital Line Graph (DLG) data, is designed to dynamically incorporate higher-resolution data or local corrections as required by users. Maintenance and enhancement will thus be shared by

data users and the USGS alike. The first data will be available online this fall; see the NHD website for the latest information.

The Reach File

With the Internet providing global electronic connectivity, increasing demand for consistency in hydrologic data has driven the joint efforts of the USGS and the EPA to produce a national hydrographic standard.

USGS's spatial hydrography layer and the Reach File's attribute model were integrated in the alpha release of the River Reach File, version 3 (RF3-alpha, or RF3). Several authors have described RF3 history and development. Dulaney (1991) described the role of the 1:100,000-scale DLG as the spatial base for RF3. Subsequently, Horn et al. (1994) detailed the history of the Reach File from its inception in the 1970's, through intermediate versions RF1 and RF2, to RF3, available since 1993. Dewald and Olsen (1994) provide a good overview of the River Reach File in a national context; the complete technical reference can be found online (EPA 1994).

In California, the River Reach File is a set of 33 INFO tables related to a set of 33 ArcInfo hydrography coverages distributed by the Stephen P. Teale GIS Technology Center (Teale). Teale's involvement in the state's hydrography data began in 1988, when USGS's DLG-3 hydrography was released, and Teale subsequently aggregated 3,200 DLG files into 33 ArcInfo coverages. This data was sent to EPA in 1992 to become the base map for California's RF3-alpha data. Both the EPA and Teale created database fields to track data changes, providing crucial documentation that enables data tied to older versions of RF3 to be linked to newer versions. This metadata is retained in the NHD.

From 1992 to 1997, the California Department of Fish and Game coordinated the development of RF3 among several organizations, including ICE. Hydrologic Unit Code boundaries, hydrologic address sequences, DS3 tables, names, and name codes were among the fields updated and corrected in California's Reach File. In 1996, changes introduced when the EPA recompiled portions of California's Reach File were also incorporated. At that time, DFG inserted the Teale primary key for hydrography, TDCKEY, as a foreign key into the tile-based tables. (August 28, 1996, is the freeze date for Teale TDCKEY.) March 19, 1997 is the freeze date for the current set of California updates to RF3-alpha. Teale forwarded the revised hydrography to EPA contractors for assembly into the NHD. A detailed history of California's experience with the Reach File is recounted in Veisze et al. (1997).

Adapting the Reach File resulted in a substantially complete ArcInfo hydrography coverage for California that is corrected for the major rivers and streams that carry most of the state's surface flow. The availability of corrected Reach Files has permitted a variety of analyses not previously possible, such as spatially analyzing regulatory options and developing tools to assess the potential impacts of upstream restoration options on downstream water quality, wildlife and fisheries.

Numerous Federal, State and local projects have used the Teale Hydrography layer as a base on which to tie information. Some examples of such projects include the following (Veisze et al. 1997):

1. California Resources Agency and ICE:
    - California Rivers Assessment: RF3 is the standard for coding locations (geocoding) of CARA data and metadata;
    - California Resources Evaluation System: internet-enabled, interagency sharing of watershed information,
    - Watershed Information Technical Service,
    - Natural Community Conservation Planning,
    - Sierra Nevada Ecosystem Study, and others.
2. Department of Fish and Game:
    - Geocoding historical information,
    - Mapping aquatic habitats with dynamic segmentation,
    - Mapping fish species ranges,
    - Managing biodiversity,
    - Fishing regulations,
    - Links to federal data, etc.
3. DFG, Wildlife Protection Division:
    - Tracking streambed Alteration Agreements ("1603 Agreements").
4. DFG, Natural Heritage Division:
    - Aquatic and riparian habitats of rare, threatened, and endangered species and species of special concern.
5. DFG, Environmental Services Division:
    - Water rights and instream flow investigations,
    - Tracking EIR/EIS documents, and
    - Streamflow estimation.
6. DFG, Inland Fisheries Division:
    - Fishery restoration sites,
    - Stream management plans
    - Fish screens, and
    - Water diversions.
7. DFG, Inland Fisheries Division North Coast Basin Planning (Eel River Basin):
    - Dynamic segmentation of stream habitat inventories,
    - Watershed problem sites, and
    - Ranges of anadromous fish species.
8. State Water Resources Control Board, EPA, and California Department of Forestry:
    - GeoWBS, water quality data spatially tied to RF3-alpha (with an Avenue program developed at ICE) to support the Clean Water Act.

The NHD

The National Hydrography Dataset is a framework dataset that combines the best of RF3 and the USGS DLG hydrography files. From RF3, the NHD benefits from hydrologic

ordering, hydrologic navigation for modeling applications, and a unique identifier (reach code) for surface water features, while the superior spatial accuracy and comprehensiveness of DLG hydrography provides the NHD with a stable spatial base.

USGS and EPA formalized their commitment to merging RF3 and DLG hydrography in a Memorandum of Understanding in 1994. The goal was to create a nationally consistent, well-documented, and dynamically maintainable hydrographic standard. This merger coincided with efforts led by the Federal Geographic Data Committee (FGDC) to develop the National Spatial Data Infrastructure (NSDI). Consequently, the National Hydrography Dataset incorporates the NSDI framework criteria set out by the FGDC (USGS 1999).

The NHD production process involved the preparation and integration of DLG and RF3, the distributed visual review and correction of this preliminary NHD product, and the subsequent central review of the completed NHD.

Data preparation and integration was an automated process called the "Blind Pass," during which the content and organization of the DLG and the RF3 were converted into feature-oriented interim datasets, known as DLG-F and RF3", respectively. These datasets were overlaid to facilitate the automated transfer of Reach File attributes onto the DLG linework. As part of the Blind Pass processing, extensive content verification was performed, including the validation of RF3 names against the latest USGS Geographic Names Information System (GNIS II). Centerlines for two-dimensional watercourses were also produced during this stage. The combined sets of relevant datasets were distributed to different regions of the country in the form of quad-based ArcInfo workspaces.

The next phase, "Visual Pass" processing, was carried out at ICE for the state of California. During this phase, a series of custom ArcInfo tools led analysts through the visual review and correction of the Blind Pass output (Figure 1). Initially, quad-based quality checking was performed to resolve conflation or centerline errors. Then the navigation data was checked within each USGS Cataloging Units (CU) to ensure the linear integrity of the reach networks.

In the final review and data-conversion phase, waterbody reaches are added to polygonal waterbodies, final GNIS II names are inserted, flow features are checked and corrected, and the data is loaded into the final database. This is the Feature Operational Database, an Oracle SDE database located at USGS's EROS center in South Dakota.

The California hydrography, updated over five years by DFG, ICE, and Teale, was accepted as California's input to the Blind Pass. The EPA Reach File work group agreed to maintain the Teale's unique identifier field, TDCKEY, during both the Blind Pass and the Visual Pass.

Transitions

The NHD presents the user community with a very different data model than previous Reach File versions. For ArcInfo users accustomed to a spatially based system of points, lines and polygons, this new framework may initially be a source of confusion. To help users understand this new structure, the NHD website provides technical documentation (summaries as well as in-depth details), sample data, a demonstration ArcView project, and a series of tutorials.

The capability to move attribute data from the RF3-alpha base to the NHD is crucial to many agencies who would like to take advantage of the enhanced features of the NHD in programs that are presently tied to the RF3.

Crosswalk tables that tracked all changes to the original RF3 reach code were maintained throughout the entire process of constructing the NHD. For instance, during the Visual Pass, the analysts changed reach code assignments for some reaches. As part of the Visual Pass processing tools, Arc Macro Language (AML) programs tracked these changes. The cross-reference table created (RCHXREF) contains a history of the evolution of each reach, starting with the RF3-alpha reach code that existed in 1994, documenting each change made and the source of the change, ultimately ending with the NHD reaches.

TDCKEY, a primary key in the Teale Hydrography layer, was tracked in separate crosswalk tables. An additional cross-reference table may be required for California users who have attribute data tied to TDCKEY to carry that data forward to NHD. This may not be a direct one-to-one relationship, and in fact may contain numerous many-to-many relationships. For example, one TDCKEY may be split and have portions in two or more NHD reaches. Conversely, several TDCKEYs may become a single NHD reach.

Until the utility and success rate of these cross-reference tables is tested, users with data attributed to the RF3-alpha base will not really know how well their data will transfer to the NHD base.

Methods

EPA's contractor, Horizon Systems Corporation, has set up two prototype cross-reference methods for the San Pablo Bay CU: tabular and spatial. (Figure 2 shows a screenshot of a sample spatial comparison in ArcView.) ICE is currently developing AML and AVENUE code to generate statistical summaries of migration errors (both commission and omission) for five selected CUs. We will be testing the success of crosswalking one-dimensional RF3-alpha arcs onto both one-dimensional and two-dimensional NHD features. (For instance, one-dimensional lake shorelines should have migrated to two-dimensional lakes; likewise for banks of double-line streams.) The next step will be to categorize types of errors and determine the stage(s) of processing during which they occur. Working with both federal and state cooperators, we will prioritize issues that require more attention and decide what steps to take in order to decrease the rate of cross-reference errors.

A crucial step will be to run the refined testing procedure using a sample data set attributed to TDCKEY (e.g., GeoWBS data). Based on the nature and source of errors encountered, ICE will recommend modifications to the cross-reference programs. We expect to work closely with EPA and their contractor to implement these changes.

Acknowledgements

We gratefully acknowledge the assistance of Tommy Dewald, EPA Office of Water; Tracey Szajgin, INDUS Corporation; Paul Veisze, DFG; Keven Roth and Tom Sturm, USGS; Terry Fleming, EPA Region IX; and Josh Viers, ICE. Furthermore, we are eternally indebted for the infinite time and resources put at our disposal by Cindy McKay of Horizon Systems Corporation.

The work described here is supported by a Framework grant from the Federal Geographic Data Committee, and includes in-kind support from:

- The California Department of Fish and Game,
- Teale Data Center, GIS Solutions Group,
- The California Department of Forestry and Fire Protection, Forest and Resource Assessment Program,
- The US Environmental Protection Agency, Region IX,
- The US Environmental Protection Agency, Office of Water,
- The US Geological Survey, National Mapping Division, Western Region, and
- The Center for Ecological Health Research, University of California Davis.

The information in this document may not necessarily reflect the views of these Agencies and no official endorsement should be inferred.

References

Dewald, T.G., and M.V. Olsen. 1994. "The EPA reach file: a national spatial data resource," in *Proceedings of the Fourteenth Annual ESRI User Conference*, Environmental Systems Research Institute, Redlands, CA.

Dulaney, R.A. 1991. "The U.S. EPA River Reach File 3: a national hydrographic database for GIS analyses," in *Proceedings of the Eleventh Annual ESRI User Conference*, Environmental Systems Research Institute, Redlands, CA.

Horn, R.C., L. McKay, and, S.A. Hanson. 1994. " History of the U.S. EPA's River Reach File: A National Hydrographic Database Available for ArcInfo Applications ," in *Proceedings of the Fourteenth Annual ESRI User Conference*, Environmental Systems Research Institute, Redlands, CA.

US Environmental Protection Agency. 1994. *The U.S. EPA Reach File Version 3.0, Alpha Release (RF3-Alpha), Technical Reference*.

US Geological Survey, 1999. *The National Hydrography Dataset*.

Veisze, P., K. Beardsley, J.F. Quinn, J. Viers, I. Oshima, M. Byrne. 1997. "California's Experience With the River Reach File," in *Proceedings of the Seventeenth Annual ESRI User Conference*, Environmental Systems Research Institute, Redlands, CA.

Author Information

Cynthia L. Moore, GIS Analyst
James C. Mullins, GIS Analyst
Karen B. Willett, GIS Coordinator
James F. Quinn, Professor
Information Center for the Environment
Department of Environmental Science and Policy
University of California Davis
One Shields Avenue
Davis, CA 95616 USA
Telephone: (530) 752-0532
Fax: (530) 752-3350
E-mail: cindy@ice.ucdavis.edu
E-mail: jcmullins@ucdavis.edu
E-mail: kbwillett@ucdavis.edu
E-mail: jfquinn@ucdavis.edu