

## **Making the Digital Water Flow**

### The Evolution of Geospatial Surfacewater Frameworks

(T.Dewald, USEPA, Office of Water, Washington, DC - June, 2015. Revised June, 2017)

The application of digital stream networks and associated catchments (local drainage areas) to support the modeling of water quantity, water quality and landscape processes continues to grow. Beginning with the Environmental Protection Agency's Reach File Version 1<sup>1</sup> in 1982, these enabling geospatial surfacewater frameworks continue to evolve and improve with the availability of increasingly detailed source data and more powerful information technology.

In the late 1970's, Robert C. Horn of the EPA's Office of Water envisioned and oversaw the development of the first national digital stream network, known as Reach File Version 1 (RF1). This initial national surfacewater framework leveraged the hydrography from U.S. Geological Survey (USGS) topographic state map series with enhancements made by NOAA (National Oceanographic and Atmospheric Administration) to deliver a 1:500,000-scale network with stream names, a stream addressing system, catchments and streamflow estimates. RF1 was then, and still is, used to support national water quality related applications, including USEPA effluent guidelines development and USGS SPARROW (Spatially Referenced Regression on Watershed Attributes) models. The local drainage areas used with RF1 have evolved over time from individual 1km elevation cells to hydrologically-conditioned 100m elevation-based catchments. Horn, known as "the father of the EPA Reach Files", also succeeded in producing a prototype RF2 product that ultimately led to the initial 1:100,000-scale RF3. The early EPA applications of the Reach Files were predominantly limited to mainframe users. However, with the growing access to desktop computing and Geographic Information Systems (GIS) technology, the Office of Water recognized the potential for widespread use as well as the associated maintenance responsibilities. In the early 1990's, Mark Olsen (EPA Office of Research and Development) and Tommy Dewald (EPA Office of Water) collaborated on a software bridge to port RF3 from the mainframe environment into desktop GIS and, also, on a prototype Reach File Update Management System (RUMS).

During the same period, the recently re-sanctioned Federal Geographic Data Committee (FGDC) initiated broad discussions that ultimately led to the pursuit of a National Spatial Data Infrastructure (NSDI). The NSDI included the development and maintenance of national framework datasets for common data themes including hydrography. As part of this activity, EPA was invited to present RF3 and RUMS to the FGDC Framework Working Group. In this setting, the EPA and USGS initiated a project in 1994 to fully integrate the 1:100,000-scale EPA RF3 stream network and names with the latest USGS 1:100,000-scale hydrography. Ultimately, this collaboration, which leveraged EPA's water applications expertise with USGS's geospatial data production and maintenance infrastructure, yielded what we now know as the medium-resolution National

Hydrography Dataset (NHD). The FGDC Subcommittee for Spatial Water Data (then chaired by Ken Lanfear, USGS Water Division), which was co-sponsored by the Advisory Committee on Water Information, promoted the development and maintenance of the NHD along with the National Elevation Dataset (NED) and the Watershed Boundary Dataset (WBD).<sup>7</sup>

The goal of the initial NHD project was to develop an application-ready, maintainable 1:100,000-scale stream network. In support of the project a Memorandum of Understanding was signed by the EPA Office of Water, the EPA Office of Information Resources Management, the USGS Water Division and the USGS National Mapping Division. Keven Roth (USGS) and Tommy Dewald (EPA) were the project co-leads. During the 1994-1996 timeframe, the multi-agency NHD design team discussed data model considerations, surfacewater feature delineation rules, production tools, stewardship concepts and conducted technology ‘science projects’, including automated conflation, centerlining through waterbodies, and stream name application from the separate USGS Geographic Names Information System. Keven Roth insisted that a surfacewater feature wasn’t legitimate until it appeared on a postcard, which she dutifully collected on one wall of the USGS conference room where the NHD team occasionally met.

In 1997, NHD production began in earnest with the automated integration of the EPA RF3 stream attributes and the USGS Digital Line Graph hydrography, a.k.a., “the blind pass”. The volume of data processing stretched the server technology of the day to its limits and occasionally beyond. For the following 3 years, a nationally distributed team of geospatial analysts from EPA, USGS and key state partners reviewed and processed the data using semi-automated GIS-based editing tools, a.k.a., “the visual pass”. Again, stretching desktop GIS software and hardware technology to its limits. Extensive quality assurance and control were applied throughout the production process including independently developed data integrity checks. Many of the lessons learned from RF3 production were applied during the NHD production effort.

In 2000, the last NHD reach was loaded into the Feature Operational Database (FOD), the NHD central repository hosted at the USGS EROS Data Center in Sioux Falls, SD. What were originally quadrangle-based, featureless “blue lines” from the 1:100,000-scale topographic maps were now watershed-based surfacewater features forming a seamless national digital stream network. The NHD data were managed and distributed by hydrologic cataloging units ... referred to by some as “cataloging unit cans”. Foreshadowing the benefits of the dynamic Internet access that we enjoy today, National Mapping Division Directory Jim Plasker observed at the time that we were going to provide NHD users with “canned green beans instead of fresh green beans!”. A good sense of humor often saved the day during the course of a long and arduous production effort.

With production completed, focus turned to maintenance with the development of state stewardship procedures and agreements that were loosely based on a “franchise concept”. In 2002, work led by the late Doug Nebert began on what would eventually become the FGDC Hydrography Framework Data Content Standard, which drew heavily on NHD content and conventions.<sup>2</sup>

In 1999, Dr. David Maidment (University of Texas at Austin) engaged ESRI in establishing the Consortium for GIS and Water Resources to help promote the broader application of the NHD, NED and WBD. In the following years, the Consortium hosted several conferences on the UT-Austin campus gathering managers and technical staff from federal and state agencies, private industry and academia to discuss progress and plans for these data and their applications. This collaboration also resulted in the development of the ESRI ArcHydro toolset. The growing interest in these events led to the establishment of what is now the biennial American Water Resources Association (AWRA) Specialty Conference on GIS and Water Resources, the first of which was held in Nashville, TN during May, 2004.

In the early 2000’s, when the USGS National Mapping Division (led by Keven Roth and Jeff Simley), the U.S. Forest Service (led by Brian Sanborn) and state cooperators initiated the production of the high resolution NHD (1:24,000-scale or better), EPA embarked on a joint effort with the USGS Water Division to develop streamflow estimates for the medium-resolution NHD. A fundamental requirement of this effort was to delineate the local drainage area (catchment) for each NHD stream segment so that ingredient data for estimating streamflow, such as precipitation and temperature, could be associated with each segment. Several different techniques for delineating catchments were evaluated, including Thiessen polygons, strictly elevation-based, and hydrologically-conditioned elevation-based.<sup>3</sup> The evaluation results showed that the hydrologically-conditioned elevation-based technique produced the best results and was feasible to implement nationally. This technique, a.k.a. the New England method, conditioned the elevation data (30m) by trenching the medium resolution NHD stream lines and raising the WBD ridgelines (for the states where they existed) in preparation for delineating catchments.

The method was developed by Rich Moore and Craig Johnston of the USGS Water Division (Manchester, NH), who joined project lead Tommy Dewald, RF3/NHD veteran EPA contractors Cindy McKay and Tim Bondelid, and USGS colleague Al Rea on the multi-agency team that would estimate streamflows for the medium-resolution NHD as part of a geospatial product suite known as NHDPlus.<sup>4</sup> Greg Schwarz of the USGS SPARROW team was a key contributor to the effort as well. Mean annual streamflow estimates were made using both the Vogel regional regressions (where appropriate) and a regression free Unit Runoff Method. In addition to the stream network, catchments and streamflow estimates, NHDPlus includes other value-added attributes that enable rapid stream network navigation, many of which drew on concepts from the original EPA Reach Files. In advance of this work, a rigorous review of the entire network was

necessary since the value-added attribute and streamflow computations required navigation of the entire country numerous times.

Like the production of the medium-resolution NHD that preceded it, the development of NHDPlus was a first-of-its-kind national effort that faced numerous challenges. The team leveraged existing tools and processes whenever possible and resorted to mailing hard drives for transporting large datasets from one member of the geographically distributed NHDPlus team to another. Many significant challenges and lessons learned are described in a National Science Foundation report released in 2009.<sup>5</sup>

Noteworthy applications of the initial NHDPlus include serving as the sample and analytical framework for EPA's National Aquatic Resource Surveys, regional SPARROW water quality models, and the Incident Command Tool for Drinking Water Protection (ICWATER). Another positive outcome enabled by the availability of NHDPlus catchments is the development of extensive collections of incremental and accumulated (upstream) landscape attributes associated with catchments.

In preparation for future streamflow estimation efforts, the NHDPlus team collaborated with the USGS Office of Surface Water in 2010 on a concept paper documenting recommended improvements to the techniques used for the initial NHDPlus streamflow estimates.<sup>6</sup> The widespread positive response to NHDPlus Version 1 is what prompted the NHDPlus team to pursue an improved NHDPlus Version 2 (NHDPlusV2) that was released in 2012.

NHDPlusV2 benefited from significant updates to the NHD network and NED elevations as well as the then recently-completed national coverage of WBD boundaries. This enhanced hydrologic foundation supported the estimation of both mean annual and mean monthly streamflows, which leveraged the recommendations noted above. The streamflow estimates were based upon runoff from the latest USGS national flow balance model refined by incremental improvements to account for excess evapotranspiration, major flow additions and removals, and gaged flow adjustments among others. The processing was conducted using a newly-developed integrated production toolset running on central servers that provided more substantial and reliable throughput, and convenient remote access for the NHDPlus team.

Additional NHDPlus applications of note include enhanced flood forecasting through the National Oceanographic and Atmospheric Administration's National Weather Service supercomputer-based National Water Model, improved access to water data through web-based services coordinated through the FGDC/ACWI Subcommittee on Spatial Water Data's Open Water Data Initiative, and the EPA-sponsored digital library of catchment-based landscape metrics, known as StreamCat. There are already, and will be, many more such applications as the water community continues to embrace, improve and apply the NHDPlus geospatial hydrologic framework.

Taking the next step in this evolution of surfacewater geospatial frameworks, during 2015 the USGS National Geospatial Program (NGP) first began pursuing a high resolution NHDPlus (NHDPlus HR), which is based upon the high resolution NHD network, 10m elevation, and the WBD boundaries. High resolution NHD is currently over 20 times as large as medium resolution NHD and, for that reason, NHDPlus HR production is applying a data development concept, based upon Agile software development, that features iterative, rapid initiation and turnaround cycles. Consistent with this strategy, the initial NHDPlus HR is being initially delivered as a beta version for review by the user community. After incorporating user-reported updates into the ingredient datasets, NHDPlus HR will be refreshed with these updated ingredients to produce a Version 1 product. This rolling update approach will produce more frequent releases and versions of NHDPlus HR.

High resolution NHD was conceived as a single 'best available' hydrography layer (1:24,000-scale or better), accommodating needs for consistent, less-detailed hydrography layers, such as 1:24,000-scale or 1:100,000-scale, through a flexible hydrologic generalization process. NHDPlus HR beta includes an initial generalization capability implemented as a visualization attribute, which can be used to extract lower resolution networks from the best-available, patchwork quilt of high resolution NHD in NHDPlus HR for purposes of display.

Looking further into the future, the USGS NGP is currently conducting a series of pilot projects evaluating techniques for generating yet more-detailed NHD networks (1:5,000-scale or better) from LiDAR (light detection and ranging) technology. Several states have already developed complete or partial coverage of LiDAR-based NHD that has been integrated into the high resolution NHD and, as a result, into NHDPlus HR. The evolution continues.

## Acknowledgements

Those who participated in these geospatial data development projects include, but are not limited to, the following:

### Reach Files

- Robert C. Horn, EPA Water (project lead)
- Tim Bondelid, EPA Water consultant
- Tommy Dewald, EPA Water
- Sue Hanson, EPA Water consultant
- Cindy McKay, EPA Water consultant

### National Hydrography Dataset (1:100,000-scale)

- Tommy Dewald, EPA Water (project co-lead)
- Keven Roth, USGS Mapping (project co-lead)
- Blaine Ailts, USGS Mapping consultant
- Duane Anderson, Pacific States Marine Fisheries Commission
- Karen Beardsley, University of California at Davis
- Tim Bondelid, EPA Water consultant
- Gladys Conaway, USGS Mapping
- Mike Domeratz, USGS Mapping
- Ellen Finelli, USGS Mapping
- Robert C. Horn, EPA Water
- Paul Kimsey, USGS Mapping
- Ken Lanfear, USGS Water
- Keith McFadden, USGS Water
- Cindy McKay, EPA Water consultant
- Sean McKinney, USFS
- Doug Nebert, USGS Water
- Mark Olsen, Minnesota Pollution Control Agency
- Bob Pierce, USGS Water
- Al Rea, USGS Water
- Bill Samuels, EPA Water consultant
- Brian Sanborn, USFS
- Jeff Simley, USGS Mapping
- Greg Snyder, USGS Mapping
- Tracey Szajgin, EPA Water consultant
- Bill Wheaton, EPA Water consultant
- Paul Wiese, USGS Mapping

### NHDPlus Versions 1 and 2 (1:100,000-scale)

- Tommy Dewald, EPA Water (project lead)
- Cindy McKay, EPA Water consultant
- Tim Bondelid, EPA Water consultant
- Al Rea, USGS Geospatial
- Craig Johnston, USGS Water
- Rich Moore, USGS Water

NHDPlus HR (1:24,000-scale or better)

- Al Rea, USGS Geospatial (project co-lead)
- Ellen Finelli, USGS Geospatial (project co-lead)
- Rebecca Anderson, USGS Geospatial
- Tatyana DiMascio, USGS Geospatial
- Ariel Doumbouya, USGS Geospatial
- Cindy McKay, USGS Geospatial consultant
- Craig Johnston, USGS Water
- Rich Moore, USGS Water
- Curtis Price, USGS Water
- And others

## References

- <sup>1</sup> “History of the U.S. EPA’s River Reach File: A National Hydrographic Database”, 1994, Robert C. Horn, et al.
- <sup>2</sup> “Geographic Information Framework Data Content Standard, Part 6: Hydrography”, FGDC-STD-014.6-2008, May 2008
- <sup>3</sup> Johnston, C.M., Dewald, T.G., Bondelid, T.R., Worstell, B.B., McKay, L.D., Rea, Alan, Moore, R.B., and Goodall, J.L., 2009, Evaluation of catchment delineation methods for the medium-resolution National Hydrography Dataset: U.S. Geological Survey Scientific Investigations Report 2009–5233, 88 p., available at <http://pubs.usgs.gov/sir/2009/5233>.
- <sup>4</sup> “Applications of the NHD at the U.S. Environmental Protection Agency”, Water Resources Impact, March 2006, Volume 8, Number 2, AWRA, pp. 5-7
- <sup>5</sup> “Advancing Strategic Science – A Spatial Data Infrastructure Road Map for the U.S. Geological Survey”, pp. 51-53, 2012, The National Academies, National Research Council
- <sup>6</sup> “Proposal for the Estimation of Streamflow Statistics for NHDPus Reaches”, February 2010, K.Ries, T.Dewald
- <sup>7</sup> “Weaving the National Hydrologic Geospatial Fabric”, T.Dewald, et al, May 2011, <https://www.epa.gov/sites/production/files/2017-01/documents/weavingthenationalhydrologicgeospatialfabric.pdf>