

## **UC Davis**

### **Information Center for the Environment Publications**

#### **Title**

California's Experience with the river reach file

#### **Permalink**

<https://escholarship.org/uc/item/7b1652hk>

#### **Authors**

Veisze, Paul

Willett, Karen Beardsley

Quinn, James F.

et al.

#### **Publication Date**

1997-02-01

**Paul Veisze, Karen Beardsley, James F. Quinn, Joshua Viers, Isaac Oshima, Michael Byrne**

# **CALIFORNIA'S EXPERIENCE WITH THE RIVER REACH FILE**

---

## **ABSTRACT**

Organized hydrographic information is vital to the success of environmental planning and management. Joint efforts from the USGS and U.S. EPA to produce a spatial base and attribute model, respectively, have resulted in an alpha release of the River Reach File, a national hydrographic standard (RF3-alpha). In California, this represents over 200,000 records of hydrographic features maintained and distributed by the state's Teale GIS Technology Center in ARC/INFO format for use by public agencies and the private sector. This paper summarizes the key events, beginning in 1992, with the California Department of Fish and Game's GIS program, which have led to the revision and correction of over 60,000 RF3 records in cooperation with federal, state, University of California at Davis (UCD), and private entities. The detail in the revisions ranges from complete hydrographic address enumeration in portions of the Eel River basin, to statewide reconciliation of USGS hydrographic names with the RF1 and RF2 names. Benefits in addition to the strong cooperative relationships identified include the enabling of Internet-based search and query capabilities by state-level programs such as the California Environmental Resources Evaluation System (CERES) and the California Rivers Assessment (CARA). Future development of RF3 will focus on empowering individuals and watershed interest groups with a robust spatial and attribute standard for hydrographic information of all types, extents, and applications.

Prepared for presentation at the 1997 ESRI ARC/INFO Users Conference, San Diego, California.

---

## **INTRODUCTION**

Increasingly, both local and national efforts to protect and restore river and riparian environments depend upon a common basis for exchanging and interrelating spatial data. At the present time, the only nationwide addressing system for rivers is the 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 (for example in the National Water Quality Assessment, NAQWA) to biodiversity (the USGS BRD Aquatic Gap Analysis Program). Local environmental assessment and restoration in California is benefitting from standardized Reach File addressing through programs such as the California Rivers Assessment, the newly proposed Watershed Initiative, the California Environmental Resource Evaluation System (CERES) and habitat restoration programs administered by the EPA, the State Water Resources Control Board, Wildlife Conservation Board, and the CALFED process.

Organized hydrographic information is vital to the success of environmental planning and management. Given the complicated qualitative and quantitative dimensions of hydrography, people tend to communicate its technologies, databases and maps being no exceptions, by specializing their nomenclatures, thesauri, and other language and logic instruments. While these specializations serve the relatively small communities of technical workers, the larger needs of society for wholistic, bioregional problem solving are hampered by these same specializations. Standards of basic location and condition must be accepted, used, and validated--not just in technical rhetoric, but actually, in the full sequence of data collection, geocoding, and networked GIS. This paper describes the experience of the California Department of Fish and Game (DFG) and the University of California at Davis (UCD) with the United States Environmental Protection Agency's (US EPA) River Reach File, version 3 (RF3-alpha). We also summarize key events, beginning in 1992, leading to the validation of over 60,000 RF3 records.

The US EPA's River Reach File is a 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](#)).

---

## **PREVIOUS WORK**

River managers have employed to varying levels of success a wide array of river addressing methods. However with the advent of the Internet providing global electronic connectivity, demand for equivalent connectivity in hydrologic data has driven the on-going, joint efforts of the USGS and the U.S. EPA to produce a national hydrographic standard. The spatial base and attribute model, from each agency respectively, are 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 Digital Line Graph (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, completed in 1982 and 1988 respectively, mapping with successively greater detail the nation's surface waters. The

current release known as RF3 has been available to a limited user community since 1993. [Dewald and Olsen \(1994\)](#) provide a good overview of the River Reach File in a national context. A brief explanation and diagram of the RF3-alpha Reach ID is presented here for readers' convenience ([Figure 1](#)). A complete technical reference for RF3-alpha can be found at <http://www.epa.gov/OWOW/NPS/rf/techref.html> (US EPA 1994).

---

## RF3 PROBLEMS

Adding the element of location to hydrologic flow models is a challenging task, compounded by the wide variety of data types that must be sampled and integrated for each element of a spatial, logical model. As a preface to working with RF3, users should also understand that DLG is a cartographic model, and its application as a base for modeling hydrologic flow brings a host of problems commonly observed when systems are adapted for purposes beyond their intended designs. While hydrographic detail such as the representation of wide rivers with left and right shorelines is useful for visualization, it tends to complicate the modeling of flow. Future designs will resolve this problem by implementing generalized centerlines or flow paths. The difficulty of linking a flow model to DLG features must have been apparent to mainframe RF3 designers, but it became even more demanding in the rigorous ARC/INFO topological environment.

Problems 1 through 9 listed below are taken from a report on the Kansas Department of Health and Environment's Surface Water Information Management System (SWIM, [Wiseman et al 1993](#)), citing other RF3 sources: Bondelid et al, Hanson et al, Howe, Kerski, and Puterski); Problems 10 and 11 are added by the authors.

1. Discontinuous network [both DLG and RF3 omission/commission errors]
2. Omitted headwater and reach segments [DLG and RF3 errors]
3. Extraneous features [knots, duplicate segments, multiple sources]
4. Generalized 1:2 million-scale HUC boundaries generating erroneous reach addresses [excessively coarse watershed boundaries, clipping errors]
5. Attribution errors [water feature names, flow direction, feature classes]
6. Inconsistent feature interpretation across USGS DLG tiles [esp. deserts]
7. Omitted cross-reference numbers from DLG source features [ok in CA]
8. Multi-date source materials used for compilation [variable map linework density]
9. Feature coordinate truncation [loss of precision in lat/long values]
10. Inconsistent implementaion of artificial flow paths through complex hydrographic linework [some open waters have centerlines, some not]
11. Lack of uniform standard for water feature names and name codes [RF3-alpha names have different formats, codes have mixed sources, including GNIS]

On a positive note, optimistic California users willing to invest local knowledge into refining RF3 have had the benefit of an emerging national hydrographic standard utilizing the best data and designs available at the time.

---

## **CHRONOLOGY OF CALIFORNIA RF3 DEVELOPMENT**

In California, the River Reach File is a set of 33 INFO tables related to a set of 33 ARC/INFO hydrography coverages distributed by the Stephen P. Teale GIS Technology Center (Teale). This section briefly describes the major events in RF3 development for California

Teale was the first California agency to address the need for a standard digital hydrographic base on a statewide scope. After the release of the DLG-3 hydrography in 1988, Teale, under the direction of Nancy Tosta, began compilation of a statewide, digital hydrography map. Over 3200 original USGS DLG files were aggregated into a more manageable set of 33 ARC/INFO coverages. Teale's tiling scheme closely matched that of the USGS 1:250,000-scale quad series (Wong-Coppin 1996 <http://www.gislab.teale.ca.gov/meta/hydrogra.txt>). The individual coverages made it possible to more efficiently process the over 220,000-record database--given the compute-intensive, graphical/topological and attribute editing involved. Edgematching, minor DLG attribute editing, and tile restructuring were conducted over approximately one and a half person-years.

In 1992, the Teale version of DLG-3 hydrography was sent upon request to the U.S. Environmental Protection Agency (US EPA) Office of Water for use as a basemap for RF3-alpha. Both Teale and US EPA staff created database fields to track data sources. In so doing, Teale joined the U.S. Geological Survey's (USGS) Water Resources Division (WRD), USGS National Mapping Division (NMD), and other states as active participants in the National Hydrography Dataset (NHD) development. Concurrently, in late 1992, the California Department of Fish and Game's (DFG) Inland Fisheries Division (IFD) obtained RF3 copies for development as a primary hydrographic base (Eric Wilson, US EPA Region IX Reach File Coordinator, 1992 personal communication).

In 1994, DFG began coordinating RF3 development among several concerns, including the University of California at Davis (UCD) and their California Rivers Assessment (CARA). CARA is an interagency program at the University of California, Davis, co-sponsored by 28 federal, state, and private resource agencies and conservation programs. Its goal is to map and assess the status of selected riparian and instream resources to assist in managing water allocations and other aspects of environmental planning. Data collection began in December 1993, with the California Resources Agency, several programs within EPA, and the National Park Service all providing substantial support (approximately \$900,000 to-date). Many more programs have provided data and

technical assistance. CARA now holds statewide and regional ARC/INFO coverages for nearly 100 themes related to river resources, land use, and conservation and restoration related projects and organizations.

Horizon Systems Corporation, US EPA's prime contractor for RF3 development, recompiled selected California river basins in late 1995/early 1996 (Lucinda McKay 1997, personal communication). This recompilation, as well as some database field type changes by US EPA rendered obsolete the tables associated with the Teale hydrography (tables designated with file extension .DS2). DFG proceeded on the information that the most current version of the River Reach File was embodied in the tables on line at the US EPA National Computer Center (NCC)(tables designated with the .DS3 extension).

DFG downloaded fresh DS3 tables from US EPA NCC in September 1996. DFG appended the entire California set, then extracted 33 tables to match the 33 tile-based, Teale hydrography ARC/INFO coverages. DFG also inserted the Teale primary key for hydrography, TDCKEY, as a foreign key into the tile-based DS3 tables. August 28, 1996 is the freeze date for Teale TDCKEY.

March 19, 1997 is freeze date for the current set of California updates to RF3-alpha. Teale has forwarded DFG- and cooperator-revised hydrography to US EPA contractors for assembly into RF3-final, described in the Future Directions section below. The Teale hydrography and DS3 tables transmittal was acknowledged on April 18, 1997.

---

## **METHODS**

Given the above problems, we proceeded on the guideline that the detail in the data to be mapped drove the detail in the level of revision applied to RF3. This ranged from intensive validation of upstream/downstream flow paths to in-stream fish habitat surveys (measured to the foot), to simpler error fixes in the set of some 150 Catalogic Unit (CU) codes for the whole of California. The spatial extent ranged from the Eel River basin for the intensive edits, to statewide for reconciliation of CUs and other USGS hydrographic names. The following sections describe the editing processes going from the simpler to the more complex.

### **Hydrologic Unit Code Boundaries**

We integrated two versions of federal Hydrologic Unit Code (HUC) coverages and two state watershed coverages to arrive at a reliable set of watershed boundaries that could serve to validate the area-based component of the RF3RCHID, the Cataloging Unit (CU). Beginning with a 1:2,000,000-scale HUC coverage (HUC2MA) obtained from the USGS Water Resources Division (WRD), Menlo Park, CA, DFG refined the boundaries to make them compatible with the Teale hydrography, i.e. so that boundaries of a watershed

would not intersect headwater streams of the hydrologic network. The DFG HUC coverage of California and neighboring states was also reconciled with the 1:250,000-scale, nationwide coverage (HUC250) informally distributed by USGS WRD out of Reston, Virginia (Doug Nebert, USGS/WRD 1995, personal communication). While HUC250 had much better precision than HUC2MA, it was still not capable of 100% accuracy in the clipping of 100K DLG Teale hydrography.

California State government applications generally use a hierarchical system of watershed designations originally developed by the Department of Water Resources. A statewide refinement of 1:500,000-scale hydrologic basin boundaries (Teale's HBASA coverage) was initiated by the California Department of Forestry and Fire Protection (CDF) by re-mapping HBASA on the 1:24,000-scale, 7.5'-minute USGS quad base. DFG joined this effort as a member of the Interagency California Watershed Mapping Committee (CALWATER) and integrated the HBASA and CALWATER coverages into the DFG HUC editing process. Necessarily, the 24K CALWATER coverage had a very high vertex density in its boundaries. This made for an ARC/INFO coverage in excess of 28 megabytes in size, with associated difficulties of transfer to and use by microcomputer GIS applications. The present DFG coverage HUCDFG1D was edited with Teale hydrography in the background, together with HUC250 and CALWATER. The objective was to create a coverage that had the benefit of very accurate clipping of 100K hydrography, without the very high boundary data density of a 24K coverage. Further, and more importantly, the DFG HUC coverage would be an input to development of a look-up table and map of state watershed boundaries nested into federal units, and vice-versa.

### **Hydrologic Address Sequences and DS3 Tables**

The RF3 Reach ID (RF3RCHID) is the primary key of the RF3 system and is incorporated in tables (designated by US EPA with .DS3 extensions) as an 18-byte character string. DFG IFD initially attempted to validate all RF3RCHIDs in the Eel River basin, to enable network analysis of streams in conjunction with detailed fish habitat databases. This involved the posting of a hardcopy USGS 7.5' quad on an easel (pre-scanned image days) alongside the monitor of an ARCEDIT session, and manually verifying the RF3 address sequence, from the mouth of a watercourse to its headwaters. As DFG neared completion of its work in the Eel Basin, UCD/CARA contracted with DFG to extend the RF3 correction work to 13 additional basins representative of hydrography across the state. The RF3RCHID sequence validation was eventually abandoned because a growing array of errors dimmed the prospects of uniform, programmable error corrections. There were in excess of 20% of the original records with addressing problems. However, the close examination of RF3 at the reach address level lent insight into what might be more practical, short-term solutions to geocoding river databases: dynamic segmentation using routes built on validated watercourse names and name codes (Byrne 1996). RF3 validation then proceeded cooperatively with DFG and UCD/CARA working on the RF3 name and name code dimensions alone.

## **Names and Name Codes**

The most basic level of hydrologic naming involves the river basin. DFG performed quality assurance on the coverage HUC2MA to reconcile its CU names with those published in "Hydrologic Unit Maps" (USGS 1987). Generic parts of names such as 'lower' and 'upper' in basin names, like Upper Yuba CU 18020125, are given explicit spatial meaning, which becomes critically important given increasing numbers of participants in watershed conservation efforts. The validated CU names and HUCs were transferred to the higher-resolution coverage HUCDFG1D for use in address corrections in RF3. With a corrected set of basin boundaries and names in place, work proceeded to treat the watercourse names and codes contained within the basins.

Water feature names are by far the most common spatial reference in use with hydrologic data. Problems arise immediately for GIS users however, when large numbers of repeat instances of feature names occur within areas of interest. To resolve this, RF3 designers implemented the Primary Name Code (PNMCD), which uniquely identifies every instance of a feature name. Thus, all the 'Deer Creeks', 'Coyote Creeks', 'Mill Creeks', etc have a unique identifier to enable reliable query and display. The problem was that a substantial percentage of RF3 records contained omission and commission errors of PNMCDs and associated Primary Names (PNAME).

Building on their experience with the DFG IFD update work, the UCD/CARA program worked with US EPA, the California State Water Resources Control Board (SWRCB), and non-governmental organizations to carry out additional RF3-alpha updates in selected CUs. Concurrently, DFG undertook a statewide validation of PNAMEs and PNMCDs that had their origin at RF1 (RF3 earlier versions RF1 and RF2 addressed increasing levels of detail in representing hydrography on a national scope; certain coding elements from each version were maintained for backward compatibility; see <http://www.epa.gov/OW/NPS/rf/esripapr.html>). Both the UCD/CARA and DFG/RF1 work followed the update guidelines below.

### **Early Rounds DFG/UCD Update Guidelines**

All California records having PNAMEs of RF1 origin were reviewed. Errors of omission and commission in PNAME and PNMCD were corrected. This often involved records of RF2 and RF3 origin as well as RF1.

1. Islands -- coded with RF1 PNAME and PNMCD.
2. Braided streams -- all arcs in the braid coded with the RF1 PNAME and PNMCD.
3. Double lined streams -- both sides of the stream coded with the RF1 PNAME and PNMCD.



4. Open Water shores -- arcs coded with the PNAME and PNMCD of the outflow stream (i.e. Shasta Lake shores coded as Sacramento R). Open Water codes not updated or modified in this pass. Streams coded with their name and code up until a double line (for a lake or reservoir) occurs. Then the outflow stream's PNAME and PNMCD are assigned.
5. Double PNMCDs for single RF1 stream -- If there are different PNMCDs for arcs with different RFORGFLAGs (RF1, RF2 or RF3), then choose the PNMCD of the RF1 stream. If RFORGFLAG is the same, then choose the downstream PNMCD and apply it to the entire stream.
6. Oxbows -- If they are already coded with the RF1 stream name and PNMCD, then leave them. If they are not named at all, leave them unnamed. If any piece of the oxbow is coded with PNAME, the arc is visited and corrected as necessary.
7. Misnamed arcs have their pname/pnmcd attributes nulled (commission)
8. Unnamed arcs which should be named are named (ommission)
9. Coastal streams coded to the mouth until the Pacific Ocean shoreline is encountered. NO Pacific Ocean shoreline should be coded with a river pname/pnmcd. Pacific Ocean code takes precedence.

### **Advanced Rounds of UCD/CARA Name and Name Code Corrections**

Increasing demands for validated stream names and name codes prompted UCD/CARA to develop more formal rules for the correction processes. The rule set, a series of IF/THEN statements, prescribing corrections to possible omission/commission errors of PNAME/PNMCD, is presented in [Appendix A](#).

### **Topological Flow**

DFG applies ARC/INFO dynamic segmentation for certain riverine data collection, mapping, and analysis. While consistent flow direction is not required in the creation of networked route/sections, it is highly desirable, particularly for mapping fish habitat. The daunting task of validating arc flow direction over the 220,000-plus arc segments in the California RF3 was made possible by developing a pathwalking algorithm which corrects the direction of each arc. The algorithm, written in Arc Macro Language (AML), processes a set of arcs with the same PNAME/PNMCD pairs (Byrne 1996). It begins processing at the headwaters of the PNAME/PNMCD set and maintains that all arcs in the set have a downstream direction. That is, the AML physically flips each arc in the set which does not already have a downstream orientation in terms of FNODE#/TNODE#. For further discussion of this process and the AML for flow corrections see <http://www.esri.com/base/common/userconf/proc96/TO250/PAP218/P218.HTM>.

Although specific discussions of the flow AML results are beyond the scope of this paper, operation of Byrne's AML yielded additional discoveries of PNAME/PNMCD inconsistencies which were corrected in later UCD/CARA quality assurance passes.

---

## **RESULTS**

Results and discussion from the RF3 work areas of HUC boundaries, RF3 addresses and DS3 tables, and naming are presented below. These results should be considered preliminary, as work is still in progress for these areas as well as for topological flow. Further, national efforts are expected to produce RF3-final by end 1997, which will supercede the present RF3-alpha implementation. Updates and corrections described here will be incorporated in the next release of RF3 (RF3-final).

### **HUC boundaries**

The entire set of RF3 CUs in California is compatible with the Teale 100K DLG hydrography and with the higher-resolution CALWATER watershed boundary coverage. There are a few exceptions where federal and state watersheds have explicitly different configurations, such as around recently created reservoirs and in areas of altered surface drainage in the Central Valley and urbanized coastal areas.

The edited CU boundaries separate drainage basins in most cases. However, they are not intended to represent true ridge lines or other topophysical features. The accuracy of the separation was greatest where the terrain strongly defines watershed basins. Accuracy was more difficult to achieve in valley areas or where administrative, rather than hydrologic, CU breaks occur (such as along a county line, water district boundary, etc). In such cases, the change in CU within the EPA RF3 address, assigned as arc attributes along watercourses, should take precedence over the CU boundary, until such time as RF3 is re-processed using the most current CU boundaries. Clipping problems are expected with CUs incorporating the Central Valley perimeter (aka "groundwater line") due to a high degree of crenulation of that boundary. Interagency review of California hydrologic basin boundaries is in progress as of this writing.

An accurate, documented CU coverage will be essential for future inter-state coordination of RF3 development, as many CUs overlap state and international borders. For California this means RF3 coordination is needed with Oregon, Nevada, Arizona, and Mexico. DFG's HUCDFG1D coverage has been incorporated into US EPA's national HUC coverage composed of refined HUC coverages from various sources (Richard Dulaney, Lockheed-Martin 1996, personal communication).

## RF3 Addresses and DS3 Tables

Some statistics on the overall California domain of RF3 are presented below, with accompanying discussions.

| <i>California RF3 Statistics</i>                   |              |
|--|--------------|
| <b>Description</b>                                 | <b>Count</b> |
| Teale/100K DLG hydrography AAT, total records      | 227,653      |
| Teale tile neatlines (non-hydro arcs), AAT records | 7,315        |
| Net Teale California hydrography AAT records       | 220,338      |

Teale hydrography coverages need neatlines enclosing the linework so that open waters such as lakes and reservoirs straddling tile edges can be properly represented as closed polygons. While edgematching within Teale tiles is very good, between-tile edgematching has not been completely validated. Users are cautioned, when undertaking wholesale removal of neatlines, to verify open-water polygon integrity (Eric Lehmer, UCD/CARA, 1997, personal communication).

| <b>Description</b>   | <b>Count</b> |
|--|--------------|
| Total California DS3 table records                         | 238,778      |
| DS3 records without links to Teale hydrography AAT records | 18,438       |
| Net California RF3 DS3 records                             | 220,340      |
| DFG-generated new DS3 records                              | 16,497       |
| DFG-generated RF3 features                                 | 0            |

The difference in the DS3 record count between total and net is due to two factors: 1) the US EPA NCC exports RF3 to ARC/INFO users on a CU by CU basis, i.e. by irregular watershed boundaries, whereas Teale maintains 33 rectangular coverage tiles, some of which extend beyond the California border. Therefore, some DS3 records from CUs overlapping California also contain records not related to the state hydrography, and 2) RF3-alpha processes had errors of omission for certain areas in the state. These were rectified by DFG by appending new DS3 records. Elsewhere, errors of commission, mostly the result of duplicate RF3RCHIDs, were corrected (rejected) by DFG by setting the TDCKEY foreign key in selected records of the DS3 tables to zero. In addition, Teale had not fully validated the precursors of TDCKEY (HYSNUM/HSCKEY) at the time of transmittal to the US EPA RF3-alpha process. As a result, some RF3RCHIDs were generated for non-hydrographic neatlines. DFG did not add or delete any spatial features to/from the Teale set. The very close match of net AAT and net DS3 record counts (220,338 vs 220,340) indicates that the California dataset is nearing a state of perfect readiness for input to RF3-final processing (described below).

| <b>Description</b>  | <b>Count</b> |
|---|--------------|
| DFG-validated RF3RCHID sequences (Eel Basin), DS3 records | 3,637        |
| Instances of replicated (non-unique) RF3RCHIDs            | 860          |
| Instances of RF3RCHID with a value of 0 for CU and SEG    | 255          |
| Instances of blank RF3RCHID                               | 15,818       |

The relatively low count of DFG-validated address sequences reflects the labor-intensive aspect of the work. Future RF3-final designs, while not likely to significantly reduce workloads for production and validation, will enable superior relational database flexibility and ease of updates and densification than would be possible with the RF3-alpha design.

The invalid RF3RCHIDs (counts of 860, 255) represent errors of commission in the Reach File, but due to the very low percentage, will be investigated in later QA/QC passes. Most records having blank RF3RCHIDs are records appended by DFG to correct RF3-alpha errors of omission with respect to the Teale hydrography. DFG and cooperators included PNAME/PNMCD and CU in appended records; RF3RCHID was left blank due to workload considerations and the expectation for US EPA corrections.

| <b>Description</b>                              | <b>Count</b> |
|---|--------------|
| Net California RF3 DS3 records (repeat)         | 220,340      |
| Number of DS3 records with non-null PNAMEs      | 63,158       |
| Number of DS3 records with non-null PNMCDs      | 144,921      |
| Number of DS3 records with null or blank PNAMEs | 175,620      |
| Number of DS3 records with null or blank PNMCDs | 75,419       |
| Number of unique PNAMEs                         | 5500         |
| Number of unique PNMCDs                         | 51,492       |

Earlier versions RF1 and RF2 had primary name codes (PNMCD) based on CU and other derivations, while name codes at version RF3 drew primarily from the USGS Geographic Names Information System (GNIS) item GNIS-ID. Future California RF3 names work will be standardized on GNIS feature names and codes whenever possible. Existing RF3 name strings entered as all-uppercase and existing non-GNIS PNMCDs will also need to be reconciled to GNIS standards.

| <b>Description</b>                                   | <b>Count</b> |
|--|--------------|
| DFG-validated PNAME/PNMCDs, DS3 records              | 55,620       |
| UCD-validated PNAME/PNMCDs, DS3 records              | 10,940       |
| California total validated PNAME/PNMCDs, DS3 records | 66,560       |
| California total validated PNAMEs                    | 2981         |

California updates to RF3-alpha as of 1997.03.19 are shown above. These results indicate that slightly more than half (2981 out of 5500, 54%) of the existing instances of PNAMEs in RF3 have been validated. In terms of records a smaller proportion of PNAMEs are verified (66,560 out of 220,340 or 30%). A significant number of records remain in need of first-time name and code assignments (175,620 and 75,419, respectively). The role of the PNMCD in RF3-alpha is not entirely clear, apart from separation of identical, common PNAMEs, because 43,916 of the 51,494 unique instances of PNMCD are associated with blank or null PNAMEs.

---

## **BENEFITS**

The progress on updating and adapting the Reach File has resulted in an ARC/INFO hydrography coverage for California that is substantially complete and corrected for the major rivers and streams that carry the majority of the state's surface flow. On a record count basis, 100% of RF1 and RF2 PNAME/PNMCDs have been validated. Approximately 7% of RF3 records have been reviewed. The extent of the validated Reach File is likely to cover the majority of available biological data (e.g. fish occurrences) and physical data (gauging stations, water quality). Nevertheless, water feature names and other attributes are still missing or uncorrected for perhaps 70% of the total river miles, mostly in low order streams.

In our own programs, the availability of corrected Reach Files have permitted a variety of analyses not previously possible. For example:

- Fish and amphibian occurrence records previously located only by coordinates, county/river name/mile or township/range/section can now be attached to specific waterways. This permits us to use connectivity data (as well as habitat data newly attached to reaches) to model where critical fish populations might be found in contiguous drainages.
- Development of a viewer and editing tools enable Regional Water Quality Control Board personnel to visually investigate landscape-scale settings and to designate the status of water bodies (e.g., querying for waterbodies fully supporting a particular use at all locations upstream of a particular confluence).
- Support for spatial analysis of regulatory options and automation of reporting required under the Clean Water Act.
- Development of tools (the "California Riparian Evaluation System") to assess the relative potential impact of various upstream restoration options on downstream water quality, wildlife and fisheries values, and other elements needed to allocate restoration funds under the Clean Water Act and recent bond initiatives.

Numerous other California programs need the corrected Reach Files. For example The California Department of Fish and Game envisions a wide variety of applications, not limited to the following:

- Resources Agency (statewide): California Rivers Assessment (CARA): RF3 is the standard for coding locations (geocoding) of CARA data and metadata; Internet-enabled, interagency sharing of watershed information: CERES, WITS, Natural Community Conservation Planning (NCCP), Sierra Nevada Ecosystem Study (SNEP), and others.
- Department of Fish and Game (statewide): geocoding historical information, mapping aquatic habitats with dynamic segmentation, mapping fish species ranges, managing biodiversity, fishing regulations, links to federal data, etc.
- Wildlife Protection Division (statewide): tracking streambed Alteration Agreements ("1603 Agreements").
- Natural Heritage Division (statewide): aquatic and riparian habitats of rare, threatened, and endangered species and species of special concern.
- Environmental Services Division (statewide): water rights and instream flow investigations; (regional): RF3 addresses will make it possible to track EIR/EIS documents and streamflow estimation.
- Inland Fisheries Division (statewide): fishery restoration sites, stream management plans (Stream Restoration Manual, Central Valley Action Plan), fish screens, and water diversions.
- DFG Region 1 (Shasta, Tehama county portions): riparian resources identified in the Sacramento River Ecosystem Protection Plan.
- IFD North Coast Basin Planning (Eel River Basin): dynamic segmentation of stream habitat inventories, watershed problem sites, and anadromous fish species ranges (raised to high priority with the recent federal listing of the Coho salmon as threatened).

General benefits to federal water programs provided by the Reach File include, but are not limited to the following (US EPA, 1986):

- Names, latitude/longitude coordinates, and other identifiers of streams and surface water features
- USGS basin codes combined with EPA reach numbers--a unified surface water classification system with foreign keys to other water databases
- Hydrologic data structures for water quality modeling and retrievals of water body data in hydrological order
- Graphical displays of water quality themes using stream, lake, reservoir, estuary, and other surface water features
- Waterbody characteristics for water quality analysis and reporting

A rich set of hydrologic routing variables as well as hydrographic features (on a national

scale) makes RF3 an ideal tool for a variety of water-related analyses. Water resource data bases maintained by federal agencies have links to RF3. Such links can provide access to analyses of water supplies, hydrology, water quality standards, and pollutant sources. A sample of EPA surface waters databases that contain RF3 links are listed below. Information in these databases is effectively mapped by the coding of specific locations along surface water features such as reservoirs, lakes, streams, wide rivers or coastlines. The following federal agencies are either using or planning to use RF3 for their special project needs (US EPA, 1993):

- USGS WRD: National Water Quality Assessment (NAWQA) studies
- USGS NMD: Future DLG releases, existing DLG updates
- National Weather Service
- Fish and Wildlife Service
- Bureau of Land Management

The following US EPA programs are utilizing RF3 (US EPA, 1993):

- Environmental Monitoring and Assessment Program (EMAP)
- Assessment and Watershed Protection Division (AWPD)
- Clean Water Act Section 305(b) Total Waters Estimates
- Waterbody System
- STORET water quality data retrievals
- Office of Information and Resource Management (OIRM)
- RF3 update procedures for ARC/INFO (EMSL)
- Hydrologic Routing

Perhaps the strongest benefit of the California River Reach File development has been, from the authors' perspective, the forging of working relationships with digital hydrography developers having national and statewide responsibilities. Specific benefits substantiating these intangibles will become apparent as validated RF3 data support increasingly reliable, Internet-based search and mapping capabilities by sites such as:

- [California Environmental Resources Evaluation System \(CERES\)](#)
- [California Watershed Information Technical System \(WITS\)](#)
- [California Rivers Assessment \(CARA\)](#).

---

## **FUTURE DIRECTIONS**

Future development of RF3 will focus on empowering individuals and watershed interest groups with a robust spatial and attribute standard for hydrographic information of all

types, extents, and applications. The US EPA and USGS are committed to further enhancing the River Reach File into a more usable, correct, and stable hydrography base layer.

Following updates and corrections to RF3-alpha, the next step toward reaching the above goals is currently under way in the compilation of the [National Hydrography Dataset \(NHD\)](#). The NHD is designed to provide comprehensive coverage of hydrologic data for the United States. NHD is based on 1:100,000-scale, Digital Line Graph (DLG) data and is designed to permit incorporation of higher-resolution data as required by users. Improved integration of hydrologically-related data is expected to serve a growing national user community and will enable shared maintenance and enhancement. As part of the dynamic nature of the National Hydrography Dataset, it will ultimately be available on-line.

The NHD presents the user community with a very different data model than previous Reach File versions. For ARC/INFO users accustomed to a spatially-based system of points, lines and polygons, this new framework may initially be a source of confusion. In order to help users of the NHD to understand this new structure, the NHD World Wide Web pages provide a [summary of the basic characteristics](#) of the new database.

The [time frame for completion](#) of the National Hydrography Dataset is an ambitious one. Processing began in the spring of 1997 and will continue through October of 1997. The first step, known as the "blind pass", will be carried out by US EPA's contractor, Horizon Systems Corporation. After the blind pass, a "visual pass" phase will be necessary in order to correct errors (i.e. conflation of attributes, flow connectivity, centerline insertion) that occurred during the blind pass. This effort will be distributed among many organizations (see the [Visual Pass Assignment map](#)), with the University of California at Davis (UCD) undertaking the visual pass processing for the State of California.

California is considered a "special case" state, as are several other states (Arizona and Pacific Northwest states) that have enhanced their state's hydrographic base maps. The US EPA and USGS have agreed to accept as input to the blind pass processing the updated California hydrography that DFG, UCD and Teale have been improving over the past five years. The US EPA Reach File work group has also agreed to maintain the Teale's unique identifier field TDCKEY (to which numerous data points have been linked) during the blind and visual pass processing. This is important because DFG and UCD have linked stream survey information and the US EPA Water Body System (WBS) to the current California Hydrography layer using TDCKEY. The UCD California Rivers Assessment (CARA) also has numerous data sets and survey results tied to the current Teale hydrography layer, as do many other agencies and organizations.

---

## CONCLUSIONS



While the relevance of nationally consistent digital databases to improved natural resource management is evidenced by ongoing deliberations aimed at establishing the National Spatial Data Infrastructure (NSDI) ([Dewald and Olsen 1994](#)), much remains to be done to develop a hydrographic framework theme sufficiently detailed and accurate to address the wide range of environmental analyses of water and river resources. The authors recognize that the Reach File, at the 1:100K DLG scale, is most useful for addressing problems in statewide and regional contexts, and is of limited applicability for local decisionmaking. This limitation however, should not obscure the continually growing need for coordination among developers and users of spatial data. Progress has been made toward such coordination with efforts like statewide watershed mapping and federated spatial datasets. However, issues of data consistency, content standards, and turf wars over data custody remain--while on-the-ground management still operates without the real benefits that are frequently promised but not fully delivered. User benefits of the River Reach file can be expected, therefore, to be proportional to level of investment in its validation.

Nevertheless, the applicability of the Reach File to resource management problems is much greater than it was just a couple of years ago. At least for the larger water bodies, significant quantities of ecological and water quality data previously held in tabular form can now be mapped and analyzed in terms of influences on and from entire drainages. Thematic coverages from disparate sources can also be inter-related, and provided, in the form of custom maps, over the Internet to users such as county planners, watershed councils, local field offices, and interested citizens. To the degree that the Reach File and similar tools catalyze information democratization effective management of natural resources will evolve from theory to reality.

---

## **ACKNOWLEDGEMENTS**

The late Tim Curtis, DFG/IFD; Virginia Wong-Coppin, Teale GIS staff; Joann Gronberg, USGS/WRD; Cheryl Henley, Eric Wilson, US EPA, Region IX; IFD GIS Staff: Mariano Arana, Jim Nordstrom, Jim Juenger; Bill O'Sullivan-Kachel, Arizona Lands Department; Doug Nebert, USGS/NMD; Mark Olsen, Tommy Dewald, US EPA/OW; Lucinda McKay, Horizon Systems Corp.; John Norton, CAL-EPA/SWRCB; CARA Staff: Kaylene Keller, Beth Kassler.

The work described here was supported by the California Resources Agency and the U.S. EPA (R819658) Center for Ecological Health Research at UC Davis. Although the information in this document has been funded in part by the United States Environmental Protection Agency, it may not necessarily reflect the views of these Agencies and no official endorsement should be inferred.

---

## APPENDIX A: California Reach File Naming Protocol and Correction Procedures

### Keywords:

RF = US EPA River Reach File 3.0 (RF3, or RF3-alpha)

reach = arc record in RF3

gnis = Geographic Names Information System

usgs = United States Geological Survey

100k = 1:100,000

24k = 1:24,000

name = placename

feat\_name = gnis field for feature name

gnis\_id = gnis field for unique primary key code

NOTE: gnis\_id, when used as a pnmcd in RF, should NOT have the leading ZERO. example : gnis\_id: 06022035 --pnmcd: 6022035

pname = RF primary name field

pnmcd = RF primary name code field

fabricated pnmcd = CCCYYMMDDNN; consists of a 3-letter EPA/NCC user id plus date (year,month,day) plus sequence number (serial number of assignment on a given day)

### Begin Rules

```
IF the reach has no pname / pnmcd pair
    THEN CHECK upstream / downstream reaches LOOK for pname/pnmcd
pair consistency
    IF reach is part of other collection
        THEN USE pname / pnmcd from that collection
    IF reach is NOT part of other collection
        THEN
            IF usgs 100k confers with gnis
                THEN CHECK 24k
                IF 24k confers with gnis
                    THEN USE gnis feat_name and gnis_id
            IF usgs 100k does NOT confer with gnis
                THEN CHECK 24k
                IF 24k confers with gnis
                    THEN USE gnis feat_name and gnis_id
            IF 24k does NOT confer with gnis
```

```

THEN USE usgs name and fabricated pnmcd
IF 24k is NOT named and gnis is named
    THEN USE gnis feat_name and gnis_id
IF 24k is NOT named AND gnis is NOT named
    THEN
        IF empirical knowledge of place name is
present
            THEN USE empirical name USE
fabricated pnmcd
        ELSE
            THEN do NOT INVOKE name/code
prodedure
IF the reach has no pname but has pnmcd
    THEN CHECK upstream / downstream reaches LOOK for pname/pnmcd
pair consistency
    IF reach is part of other collection
        THEN USE pname / pnmcd from that collection
    IF reach not part of other collection
        THEN
            IF usgs 100k confers with gnis
                THEN CHECK 24k
                IF 24k confers with gnis
                    THEN USE gnis feat_name and gnis_id
            IF usgs 100k does NOT confer with gnis
                THEN CHECK 24k
                IF 24k confers with gnis
                    THEN USE gnis feat_name and gnis_id
            IF 24k confers with gnis except minor variation
(hollow vs. holler)
                THEN USE usgs name and gnis_id
            IF 24k does NOT confer with gnis
                THEN USE usgs name USE existing pnmcd
            IF 24k is NOT named and gnis is named
                THEN USE gnis feat_name and gnis_id
            IF usgs 24k is NOT named AND gnis is NOT named
                THEN
                    IF empirical knowledge of place name is
present
                        THEN USE empirical name USE
existing pnmcd
                    ELSE
                        THEN do NOT INVOKE name/code
prodedure
IF the reach has pname / pnmcd, but they are wrong
    THEN CHECK upstream / downstream reaches LOOK for pname/pnmcd
pair consistency
    IF reach is part of other collection
        THEN USE pname / pnmcd from that collection
    IF reach is NOT part of other collection
        THEN
            IF usgs 100k confers with gnis
                THEN CHECK 24k
                IF 24k confers with gnis
                    THEN USE gnis feat_name and gnis_id
            IF usgs 100k does NOT confer with gnis
                THEN CHECK 24k
                IF 24k confers with gnis

```

```

THEN USE gnis feat_name and gnis_id
IF 24k confers with gnis except minor variation
(hollow vs. holler)
THEN USE usgs name and gnis_id
IF 24k does NOT confer with gnis
THEN USE usgs name USE existing pnmcd
IF 24k is NOT named and gnis is named
THEN USE gnis feat_name and gnis_id
IF usgs 24k is NOT named AND gnis is NOT named
THEN
IF empirical knowledge of place name is
present
THEN USE empirical name USE
existing pnmcd
ELSE
THEN do NOT INVOKE name/code
prodedure

```

## End Rules

---

## REFERENCES

- Bondelid, T.R., S.A. Hanson, and P.L. Taylor. 1990. *Technical description of the reach file* US EPA Internal document. Horizon Systems Corporation.
- Byrne, M. 1996. *California salmonid habitat inventory: a dynamic segmentation application*. in Proceedings of the Sixteenth Annual ESRI User Conference. Environmental Systems Research Institute, Redlands, California.
- Dewald, T.G., and M.V. Olsen. 1994. *The EPA reach file: a national spatial data resource* US EPA Internet document.
- Dulaney, Richard A. 1991. *The U.S. EPA River Reach File 3: A National Hydrographic Database for GIS Analyses*. Proceedings of the Eleventh Annual ESRI User Conference, Redlands, CA.
- Hanson, S.A., B. Deffenbaugh, C.R. Horn, and L. McKay. 1992. *River reach file (RF3), update and quality control, standards, procedures, and management*. Internal US EPA document. Horizon Systems Corporation.
- Horn, R.C., McKay, L., and Hanson, S.A. 1994. *The History of the Reach File*. in Proceedings of the Fourteenth Annual ESRI User Conference. Environmental Systems Research Institute, Redlands, CA.
- Howe, B. 1993. St. Johns River Water Management District, Palatka, Florida. Interview. 21 January 1993.

- Kerski, J.J. 1992. *Hydrologic analysis: DLG, TIGER, or RF3?* Unpublished paper.
- Puterski, R. 1992. *Using dynamic segmentation with the Reach File 3 stream database.* in Proceedings of the Twelfth Annual ESRI User Conference. Environmental Systems Research Institute, Redlands, California.
- US Environmental Protection Agency. 1986. *Reach File Manual.* US EPA Internal documentation of Reach File version 2 (RF2).
- US Environmental Protection Agency. 1993. *Technical Description of the Reach File.* US EPA Draft, Horizon Systems Corporation, February 1993.
- US Geological Survey. 1978. *Hydrologic unit map (California).* Water Resources Council. 1:500,000-scale USGS planimetric base, 2 sheets.
- US Geological Survey. 1987. *Hydrologic unit maps.* Water Supply Paper 2294.
- Wiseman, R.,A.J. Thomas, R.D. Miller, and M.K. Butler. 1993. Surface waters information management system. in Proceedings of the Thirteenth Annual ESRI User Conference. Environmental Systems Research Institute, Redlands, California.
- 

## **AUTHOR INFORMATION**

Paul Veisze, Spatial Data Coordinator  
Isaac Oshima, GIS Analyst  
Michael Byrne, GIS Analyst  
California Department of Fish and Game  
1730 "I" Street, Suite 100  
Sacramento, California 95814  
Telephone: (916) 323-1667  
Fax: (916) 323-1431  
E-mail: pveisze@dfg.ca.gov  
E-mail: iosshima@hq.dfg.ca.gov  
E-mail: mbyrne@dfg.ca.gov

Karen Beardsley, GIS Coordinator  
James F. Quinn, Professor  
Joshua Viers, GIS Analyst  
Division of Environmental Studies  
University of California, Davis  
Davis, California 95616  
Telephone: (916) 752-4389  
Fax: (916) 752-3350  
E-mail: kbeardsley@ucdavis.edu

E-mail: [jfquinn@ucdavis.edu](mailto:jfquinn@ucdavis.edu)

E-mail: [jhviers@ucdavis.edu](mailto:jhviers@ucdavis.edu)

---

---