History of the U.S. EPA's River Reach File: A National Hydrographic Database Available for ARC/INFO Applications

Prepared Under Contract for:

Office of Wetlands, Oceans, and Watersheds
Office of Water
U.S. Environmental Protection Agency
Washington, DC

May 25, 1994

The Authors:

C. Robert Horn
Computer Data Packet Company
42 Sixth Street
Apalachicola, FL 32320
(904) 653-9009
Sue Ann Hanson, Director of Scientific Applications
and
Lucinda D. McKay, President
Horizon Systems Corporation
423 Carlisle Drive
Herndon, VA 22070
(703) 471-0480

Introduction

The U.S. Environmental Protection Agency's (EPA) 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.

The Reach File was conceived in the 1970s with a proof-of-concept file, known as RF1A, completed in 1975. The first full implementation, referred to as RF1, was completed in 1982. RF1 was created from scanned 1:500,000 scale National Oceanic and Atmospheric Administration (NOAA) maps and consisted of approximately 68,000 reach segments comprising 650,000 miles of stream. While RF1 still supports broad-based national applications, it served as the only reach file until 1987/88. At that time, the Feature File of the U.S. Geological Survey's (USGS) Geographic Names Information System (GNIS) was used to add one new level of reach segments to RF1. The resulting file, known as RF2, contained 170,000 stream segments. Development of the current Reach File, RF3, began in 1989. The RF3 production process involved the overlay of the RF2 file, the GNIS Feature File, and the 1988 Digital Line Graph, Version 3, 1:100,000 scale, Optional format (DLG3) file. RF3 contains over 3,100,000 individual hydrographic features (reaches) and over 93,000,000 coordinate points.

A key characteristic of the Reach File is its attributes which define the connected stream network. These attributes provide connectivity regardless of the presence or absence of topologic continuity. An additional benefit of the Reach File is the flow direction inherent in the connectivity attributes. The attribute-level connectivity enables the reach file to provide hydrologic ordering (what is upstream and downstream of a given point in the stream network) as well as network navigation proceeding either upstream or downstream.

RF3 has been available to a limited user community since early in 1993. Several EPA programs have utilized the River Reach File in ARC/INFO on UNIX workstations, in custom software on the EPA IBM mainframe, and in ArcVIEW and custom software on Personal Computers. The unique reach code assigned to each reach has been used to link a number of EPA national databases to surface waters, for example STORET Water Quality, Facility Discharges, and Drinking Water Intakes. Any site within these linked databases can be associated with a specific location on a surface water feature: reservoir, lake, stream, wide river, coast line or other feature. The Reach File has also been used by the USGS and NOAA as the hydrography backbone for several of their programs and applications. The reach code is presently being considered by the Subcommittee on Spatial Water Data of the Federal Geographic Data Committee (FGDC) as a candidate for standardizing surface water location identifiers.

During 1993 and early 1994, user feedback, particularly from the portion of the user community who uses RF3 in ARC/INFO, has defined a number of desirable improvements in the Reach File. This coupled with the joint efforts of EPA and USGS to coordinate the Reach File with the new Digital Line Graph-Enhanced (DLG-E) file has resulted in a project currently underway to improve the Reach File and to synchronize it with DLG-E.

Background

The Reach File is a geographic database of surface-water features of the United States. For each feature, the Reach File database provides a unique identifier, a variety of attributes and a latitude/longitude vector trace.

Some features in the Reach File are hydrologically connected to each other forming the structure that represents the branching patterns of surface water drainage systems. Hydrologic connectivity is stored by referencing reaches to each other using their unique identifiers. Reach-to-reach referencing within the drainage system facilitates upstream and downstream traversals and mathematical modeling of hydrologic flow. Reach-to-reach referencing between non-drainage system features such as shorelines, permits circumnavigation around the perimeters of reservoirs, lakes, and estuaries and facilitates geometric analyses and displays of polygon representations of open waters.

One of the primary objectives for developing the Reach File was to provide a common geographic database for interrelating data contained in many environmentally significant databases. With the reach numbers and other data provided by the Reach File, coupled with computer generated site marker indices inserted in other files, it is possible to co-analyze data occurring in many files, and to enrich that analysis with upstream and downstream meaning. For example, pollutant discharges described in one file may be related to downstream domestic water supply intakes in another file, and these analyses may be related to data from upstream and downstream water quality monitoring stations found in yet a third file. Prior to the existence of the Reach File, it was impossible to co-analyze these and other data because they were collected by separate agencies and organizations without a common geographic framework.

In addition to data integration among diverse databases and providing downstream/upstream navigation through those databases, the Reach File provides hydrologically ordered retrievals from those databases. Hydrologically ordered retrievals from multiple water resource files are ideally suited for water quality modeling because they can and do provide needed data in the same sequence required by the models as the simulation progresses downstream throughout the dendritic systems of watersheds from points of origin to watershed outlets.

Development of the Reach File

The Reach File has progressed through four versions: RF1A, RF1, RF2 and RF3. RF1A used large natural watersheds as the primary unit for subdividing the surface water systems of the nation. RF1 and subsequent versions use the standard USGS Catalog Units for this purpose. In all, there are 2,111 catalog units in the contiguous United States. RF1A contained only a few thousand reaches, whereas there are approximately 70,000 reaches, 140,000 reaches, and 3,200,00 reaches in RF1, RF2 and RF3 respectively. Figures 1, 2, and 3 illustrate the relative reach density between the RF1, RF2, and RF3 files for a single catalog unit 02070010, representing a part of the Potomac River in the vicinity of Washington, DC.

RF1A - History And Technical Approach

RF1A was designed in 1973, funded for development in 1974, and completed in 1975. It was digitized by EPA from stable base acetate copies of the two-part U.S. Geological Survey (USGS) wall map that has a scale of 1:2,500,000. This first version of the reach file was a single national coverage used for database design testing and demonstration. It used natural watersheds which terminate at the confluences of large streams and at estuarine outlets. Its design was found to be well-suited for the integration of surface water databases and for hydrologically ordered retrievals of any data that might be indexed to it along river courses or through open water bodies.

Spatial data for RF1A were captured as latitude and longitude coordinate vectors representing the traces of stream reaches and shoreline reaches. The topological relationships among reaches were derived from the digital traces using largely user assisted interactive graphics software prepared specifically for the project. Each reach was assigned a unique reach number to be used as a reach identifier. These numbers were assigned without regard to hydrologic order because it was felt it would be unwise to burden the identifier with meaning that would likely need to be compromised to accommodate additional reaches within the database in the future. The reach identifiers were unique numbers within large watersheds. There were several hundred large watersheds covering the contiguous U.S. and Alaska which ranged in size from less than 100 reaches to several hundred reaches. Globally, the reach number used in RF1A consisted of a nationally unique 6-digit major watershed number and a 6-digit reach number unique within the watershed.

RF1A briefings were presented at EPA to the Office of Water and the Office of Enforcement, and at the U.S. Geological Survey for the Office of Water Data Coordination and the National Mapping Division. Technical presentations were also made for the Corps of Engineers, the Department of Agriculture and the Fish and Wildlife Service. Comments were solicited and received. The EPA comments pertained to water quality monitoring and reporting requirements of Sections 104, 106, and 305 of the Clean Water Act as well as the Office of Enforcement's concern to expand the database to handle more detail for what was then conceived as the BPT (Best Practicable Treatment) permitting process. The Office of Water Data Coordination requested that the Hydrologic Unit Coding System be incorporated in the reach number for the prototype database, and the Fish and Wildlife Service requested that the reach numbers be sequenced in hydrologic order. The Corps of Engineers was interested in indexing their river mile points to the Reach File on a pilot basis when and if EPA increased the detail in the database to a scale of at least 1:250,000.

RF1 - History and Technical Approach

In 1978, work began to build the Reach File into a fully functional database on the EPA's IBM mainframe computer. This work was undertaken in support of the EPA's effluent guidelines development process in the Office of Water. This was to be done in a manner which would be integrated with STORET and yet be exportable to other agencies to begin enhancing water data integration on both a national and local scale. The resulting Reach File was completed as RF1 in 1982.

RF1 was digitized by EPA from aeronautical charts prepared by the National Oceanographic and Atmospheric Administration (NOAA). A primary objective of the NOAA charts was to provide the best hydrography available in a single map series so that it could serve low altitude aircraft navigation. NOAA initially developed these charts from photographic copies of stable-base scribe coat masters of the USGS maps which have a scale of 1:250,000. Prior to the development of these aeronautical charts, the USGS 1:250,000 scale maps held the most up-to-date and detailed national hydrographic coverage of any national map series.

Although the NOAA charts retained all the hydrography shown on the USGS 1:250,000 scale maps, there was a need to improve these coverages because there was a significant disparity in the density of stream coverage from one USGS map to the next. Therefore, NOAA updated their copies of these maps to include more streams and waterbodies. The updates were based on remote sensing imagery and emphasis was placed on providing more hydrographic features on those maps which were "under par" relative to adjacent maps. This provided a more balanced hydrographic coverage for the nation as a whole than was available from any other source.

Additionally, NOAA updated these maps on a continuing basis so that they would be current with changes, such as the construction of new reservoirs in many areas of the nation.

The NOAA aeronautical charts provided, by far, the most detailed complete national hydrographic coverage of any set of maps of the nation. They were photographically reduced to a scale of 1:500,000 yet they retained all of the USGS 1:250,000 scale hydrography and all of the NOAA hydrographic updates. Therefore, they became the maps of choice for the Reach file. They were optically scanned (to eliminate manual digitization errors), edge-matched, and transformed into a single contiguous vector database by EPA during the period of 1978-1982.

Vectorization was accomplished using interactive graphics software developed for the project. This software automatically followed lines in the raster database until nodes were encountered, whereupon it paused, waiting for the operator to coax the vectorization process in the necessary direction through a few pixels, to yet again progress automatically and visually along the next stream or shoreline segment until encountering another node. Flow direction was imparted to the database at this stage, names were added later, and the data elements which explicitly define the hydrologic connectivity were populated by a subsequent, and largely automated process using other software developed for the project.

The resulting database, RF1, includes information on single line streams, a few double line wide rivers, a number of lakes and reservoirs and most estuaries. All of these features were represented by lineal constructs comprising either hydraulic transport reaches or shoreline reaches. Those representing hydraulic transport paths were referred to as transport reaches, whereas those representing the banks of wide rivers, lakes, and estuaries were referred to as shoreline reaches. Aerial features such as lakes, reservoirs, some wide rivers, and estuaries were represented by shoreline reaches and artificially constructed transport reaches which ran through the aerial reach.

RF1 - Applications

In 1983, the potential analytical power that the Reach File was to bring to water quality analysis and hydrological analysis was demonstrated to EPA and USGS using a Z80-based microcomputer. Low-end performance by today's standards, this computer had 2 megabytes of random access memory, a 20 megabyte hard disk drive, and a DEC VT120 monitor for map display. These RF1 demonstrations showed streamflow routing, water quality modeling, relocating of point sources under cursor control, and upstream and downstream traversals of reaches for analyses of point sources, water quality monitoring stations, drinking water intakes, and streamflow gages for all of Accounting Unit 030300 in North Carolina. The demonstrations used dBase-I (a relational database system) and MDBS-III (a very powerful relational database system which included the ability to represent many-to-many relationships).

Through the STORET system, RF1 went into use in support of effluent guidelines development in 1982. Reach retrieval capability was ultimately made available for the STORET Water Quality File as a result of this initiative in support of the Clean Water Act (CWA). As the STORET application was being prepared, RF1 was adopted for use by the EPA Municipal Construction Grants Program to help in the assessment of needs and priorities for municipal facilities and by the Water Monitoring program to help support the development of State reports to Congress under Section 305(b) of the Clean Water Act. A short time later, the EPA Office of Water began developing the Reach Characteristics File for the construction grants priority and needs activity and the Waterbody File for support to the Section 305(b) activity. Final development of the Reach Characteristics File was postponed, but the Waterbody File is playing a significant role in 305(b) reporting by providing for the characterization of water quality by waterbodies which states define in terms of the Reach File.

In 1984/85, the EPA Municipal Construction Grants Program used RF1 as the basis of an innovative nationwide water quality modeling application. The primary objective of this effort was to assess the relative in-stream water quality improvements that could be achieved through building and/or upgrading municipal sewerage treatment plants. Using the USGS Digital Elevation Model and the USGS Daily Stream Flow data, flows were estimated for each reach in RF1. Using the STORET Water Quality Monitoring data, background water quality was estimated for headwater reaches. The water quality modeling software used the reach navigation attributes to route flow, dissolved oxygen, total ammonia, toxic ammonia and biological oxygen demand, collecting effluents from municipal and industrial discharges, as the modeling process moved downstream through the drainage systems of RF1.

Several briefings and presentations about this new version of the database were given to the interagency community. Most notably, the U.S. Fish and Wildlife Service adopted RF1 as their standard stream identification system. USFWS prepared reference manuals, training manuals, and video presentations describing the Reach File for statewide biological surveys. The USFWS program was implemented in cooperative efforts with several states. Later, the Bonneville Power Administration also adopted the Reach File to support decision making in their environmental and power generation missions and began updating it soon thereafter.

RF1 is still in use today for many applications. Despite the recent development of the 1:100,000 scale RF3 and the potential for a 1:24,000 scale Reach File, it is believed that there are many applications at the national level that will continue to need RF1.

RF2 - An Intermediate Product

An intermediate update (RF2) was constructed during the late 1980's to aid in building a much more detailed version (RF3) that would be based on the then upcoming USGS 1:100,000 Scale Digital Line Graph (DLG) database. RF2 was developed by overlaying, onto RF1, the coordinates of the hydrographic feature names extracted from the USGS Geographic Names Information System (GNIS), Version I.

The overlay software was custom-built for the project on the EPA IBM mainframe computer. The software ran in batch mode and included many specialized functions to improve processing speed and to support quality control of the overlay process. Entering the overlay, features were screened from the process based on their estimated length thereby eliminating much overlay clutter. The software provided extensive diagnostics on the relative significance of various matches of end points and intermediate points of GNIS features with RF1 features. The simplest types of matches were those that were found to be unambiguous. These clear-cut matches resulted when all of the GNIS coordinates for a particular feature were successfully match to an RF1 stream. The overall process involved a wide variety of much more complex matches than these. Once the matching features were eliminated from GNIS, the remaining features were overlaid to find those streams which discharged into RF1 streams. These stream were added to the Reach File dataset and diagnostics for these matches were also generated by the software.

The diagnostics generated by the batch overlay process were input to interactive graphics software which was developed to validate the overlay and to provide quality control in the project. The interactive graphics were solely for quality control and no update capability was included.

The primary improvement to RF1 as a result of the GNIS overlay process was the addition of one new level of reaches and their names. Reaches were added to RF1 only when they discharged directly into an original RF1 reach. However, even more valuable than the improved density in RF2 were the lessons learned about the behavior of spatial datasets with disparate coordinate sources in what was essentially a blind overlay operation.

The RF2 development project essentially doubled the number of reaches in the Reach File, but the much awaited digital traces from the 100,000 Scale DLG database were not available at the time RF2 was scheduled for completion. Consequently, RF2 was released, in 1988, with the sparse coordinates obtained from GNIS. Nevertheless, RF2 was a very valuable Reach File for many water quality reporting needs within EPA. From 1988 until RF3's initial release in 1992, the 305(b) program used RF2 to standardize the reporting on state priority waterbodies.

RF3 - The Next Generation

The successful applications of RF1 and RF2 created the need and desire for a more comprehensive hydrologic database. To satisfy this need, the RF3 development project was begun in the fall of 1988 when the 1:100,000 scale DLG data became available. RF3 was developed by EPA's Office of Water to provide a nationally consistent database to promote comparability for national, regional, and state reporting requirements such as those found in 305(b) and other sections of the Clean Water Act.

The specific goals for RF3 were to:

- Create a nationally consistent hydrographic database;
- Assign unique, permanent reach code identifiers to all surface water features,
- Allow hydrologic ordering of reaches;
- Include spatial data for realistic representation and geo-referencing capabilities;
- Incorporate features and navigation attributes needed for one-dimensional modeling;
- Provide the means to expand to two-dimensional modeling; and
- Retain continued compatibility with RF1 and RF2.

The RF3 compilation project had limited resources and was designed around the following requirements:

Data Management Perspective:

- Provide National Access
- Accessible to Managers, as well as, Analysts
- Minimize Mainframe Storage Requirements
- Include Mainframe Graphics Capabilities
- Consider Both National and Local Reporting Requirements

Cost Perspective:

- Minimize Cost By Using Best Available Digital Source Data
- Incorporate Important Hydrologic Features From the USGS DLG
- Striving for an Overall Success Rate Of 90-95% Inclusion of DLG
- Maximize Incremental Production Capabilities to Permit Compilation, as Resources Permitted
- Minimize End-User Costs

RF3 is now complete for 45 of the 48 contiguous states and Hawaii. In its present form, RF3 includes approximately 3,200,000 reaches representing streams, wide rivers, reservoirs, lakes, a variety of miscellaneous hydrographic features, and the coastal shorelines for the Atlantic and Pacific Oceans, the Great Lakes, the Gulf of Mexico and the Hawaiian Islands.

RF3 - Technical Approach

RF3 was constructed from four different data sources:

1. EPA Reach File, version 2 (RF2)
2. USGS 1:2,000,000 Catalog Unit (CU) boundaries
3. USGS Geographic Names Information System Version I (GNIS)
4. USGS 1:100,000 scale Digital Line Graph, Version 3, Optional format (DLG3-O)

The DLG data were received from USGS in approximately 54,000 files on 241 9-track magnetic tapes. These files, one for each 7.5-minute quadrangle, were consolidated into a single contiguous file on EPA's IBM mainframe computer. Software was developed for processing this very large database in preparation for overlaying it with RF2 and GNIS. Two basic steps were performed to prepare the DLG data. The first step concatenated DLG lines when their end points were coincident and their attributes were the same. The second step combined DLG lines which were within a given tolerance. At map edges, an initial tolerance of .0003 degrees was applied, then .0006, .0012, .and finally 0024 until one and only one other DLG line was found within the search radius. If at any point more than one other DLG line was found, the concatenation was deferred to a later interactive step. Within a map, the tolerances of .0003, .0006, and .0012 were used. As DLG lines were concatenated, the order of DLG coordinates was changed, as needed, to ensure that the trace for the new, longer DLG line progressed consistently from one end to the other.

Custom software was created to perform the task of actually compiling RF3. The PC Reach File (PCRF) program was used as the primary tool for performing the networking of DLG, the overlay of DLG, RF2, and GNIS and the final construction of RF3 reach records. The capabilities of PCRF included the following:

- On-screen interactive extension of the DLG lines, as needed, to create topologic connectivity;
- Techniques to speed up display handling;
- Automated networking of DLG data (including rule-based expert system procedures);
- Transaction logs the editing and overly procedures with re-play capabilities;
- Automated overlay of RF2 Features;
- Automated overlay of GNIS names; and
- Upload/download using 3270 data transfer techniques.

The RF3 compilation was performed on a catalog unit basis. The DLG and GNIS files were divided into CU-based subsets using the 1:2,000,000 Catalog Unit Boundaries. Since the actually compilation of RF3 data for each CU was performed on a PC, the data for one CU from each of the four source files was downloaded to the compilation PC. This process was streamlined and data compaction techniques were implemented to improve efficiency. Automated data transfer procedures using IBM 3270 bisynchronous protocols greatly reduce transfer time and transfer errors.

On the compilation PC, the first step was to network the DLG. This process put the DLG lines in a hydrologic network, assigned temporary "reach" identifiers and built a preliminary set of navigation attributes. After the downstream start point(s) were selected by a technician, PCRF used an automated endpoint-to-endpoint method to find and connect all possible DLG lines into the network. DLG attribute codes were used to distinguish between single line streams, wide rivers, and lakes causing circumnavigation of open water bodies when they were encountered. Miscellaneous hydrography such as point features and ditches and canals were not networked, however, most of these features were included in RF3 and given reach codes. PCRF graphically displayed the networking as it took place so that problems could be spotted and corrected through manual intervention by a trained technician. The types of manual intervention included:

1. Marking divergent junctions,
2. Extending DLG endpoints to properly connect the network,
3. Breaking a DLG line if an obvious connection should exist but didn't, and
4. Overriding some DLG line attributes, especially mis-coded line types, to permit networking to continue successfully.

Networking the DLG was the most labor-intensive part of the RF3 compilation process. PCRF was continually made "smarter" to reduce the need for manual intervention. When a situation that required manual intervention could be translated into a set of specific rules, a networking algorithm was designed and included in the PCRF automated process. The final version of PCRF was version 8 and, by the middle of the compilation effort, the software was able to correctly identify and resolve between 97 and 99 percent of the networking problems.

The next step in compilation was to overlay the RF2 onto the DLG network and transfer the essential RF2 data elements to the DLG network. These data elements included the reach code, reach names and name codes, stream levels, and navigation attributes. The inter-CU reach connectivity was of particular importance because all of RF3's inter-CU connections originated from RF1/RF2.

After the RF2 overlay, the GNIS file was overlaid to add names for the new RF3 reaches. This was similar to the RF2 overlay, except that it was performed on a named feature-by-named feature basis. This process was streamlined to permit automated overlay techniques. The names assigned to reaches in RF1 and RF2 were not changed during this overlay. Only reaches that were newly added by DLG hydrography were candidates to receive names in this process.

The final step in RF3 compilation consisted of a comprehensive set of QA/QC procedures. Manual intervention during QA/QC was made easier with display options and diagnostic RF3 reports. After the QA/QC was complete, the final RF3 reach records for the CU were transferred back to the EPA NCC mainframe and consolidated into the national RF3 database.

RF3 - Draft Applications

RF3 was completed in its draft form in early 1993. Beginning in the fall of 1992, parts of the dataset were released to a limited user group for evaluation and pilot applications. At that time, several national EPA programs began using RF3 to provide hydrographic information for their specific program needs. These programs included:

Environmental Monitoring and Assessment Program (EMAP)

- Determine Lake and Stream Frames

Assessment and Watershed Protection Division (AWPD)

- Total Waters Estimates
- Waterbody System and 305(b) Waterbody Indexing
- Linking Environmental Information Systems
- STORET, IFD, GAGES, DRINKS - indexing to RF3
- STORET data retrievals
- STORET Modernization incorporating RF3 indexing and retrievals

Other Federal agencies have used it on a limited basis for their internal applications. US Geological Survey has used RF3 in their NAWQA Studies and for Hydrologic Modeling. RF3 is providing the geographic dataset for The National Weather Service's Flash Flood Prediction Model. The Bureau of Land Management and other federal and state agencies are currently creating RF3 for Alaska.

RF3 - Current Activities

Based on this extensive review by users inside and outside of EPA, a "To Do" list for RF3 was established. Some of the early users were GIS'ers and their requirements were very different from the original objectives, so the "To Do" list grew substantially in response to the new objectives introduced by GIS applications. As a result of this user feedback, the second part of the RF3 production process was initiated and called Validation. Validation contains two phases: Assessment and Revision. The Assessment Phase consists of gathering "hard" documentation of user feedback and performing many automated QA/QC checks on the data. The first objective of Assessment is to learn as much as possible about the problems in the data and develop approaches to fix them. Knowing that there are some problems that will not be fixable due to lack of local knowledge, the second objective is to create documentation datasets for each CU that contain CU-level and feature-level information about the data content and can be served as "metadata" along with the spatial data. The results of the Assessment Phase will establish the activities that will be performed during the Revision Phase.

Anticipating that substantial changes to RF3 will occur in the course of addressing the "To Do" list, the Assessment Phase presents an opportune time to also assess the design of the RF3 feature representation and data structures. The current design was primarily inherited from RF1 and is in need of changes to meet new requirements and to operate under recent advances in GIS technology. A new RF3 design for feature representation and data structures is currently under development.

The Assessment Phase contains fours steps:

1. Gather written documentation from RF3 initial users.
2. Perform an automated QA/QC pass to check for validity and consistency on all RF3 attributes.
3. Perform a second automated QA/QC pass to navigate the entire file searching for irregularities in the navigation attributes.
4. Perform a third automated QA/QC pass to assess topologic connectivity and to check reach-to-CU assignments against the 1:250,000 CU Boundaries.

To date the first two steps have been completed with step 2 beginning the process of creating the CU-level documentation datasets. The following table presents some of the quantitative results from step 2. Figure 4 shows the distribution of reaches by type.

RF3 General Assessment Statistics

Description	Count
Total No. of CUs	1,925
Total No. of Reaches	3,177,599
Connected	2,181,789
Start Reaches	864,929
Unconnected Reaches	995,810
RF1 Reaches	552,005
RF2 Reaches	168,662

RF3 Reaches	2,456,919
Blank Names	
RF1 Reaches	157
RF2 Reaches	74
RF3 Reaches	2,271,918

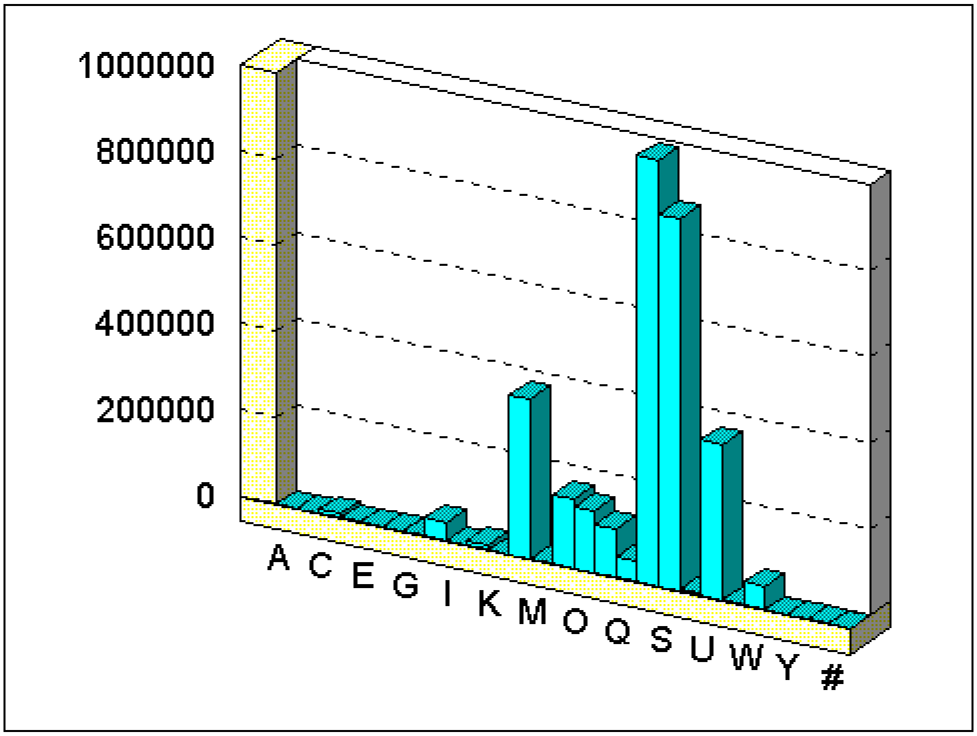

Figure 4 - Number of Reaches by Reach Type

RF3 - The Future

The immediate future of RF3 is completion of the comprehensive Revision Phase followed by the official release of the final RF3 dataset. The Revision Phase will consist of three passes of the RF3 data. Pass 1 will be a blind mainframe process which will fix various data problems and restructure the file based on a new design for feature representation and data structures. Some of the changes envisioned in this pass are:

- Re-delineate the wide rivers into confluence-to-confluence reaches,
- Add junctions to accommodate navigation attributes,
- Correct the CU Boundaries for the Great Lakes CUs,
- Possibly include Puerto Rico,
- Possibly add centerlines,

- Possibly add lake centroids as label points,
- Move reaches from one CU to another where the 1:2mil CU boundaries made an incorrect CU assignments,
- Incorporate the Pacific Northwest River Reach File,
- Add open water polygon reaches,
- Eliminate Zero-length connectivity reaches, and
- Possibly update RF3 names with GNIS II.

Pass 2 which will consist of GIS operations which will begin with a variety of "blind" GIS actions on the data, including the replacement of the RF3 linework with the latest DLG linework and adding all DLG attributes to RF3. Finally, the "blind" pass will compose a workspace ready for a visual GIS pass. The visual GIS activities will perform thorough QA/QC of the work performed in the mainframe step and in the GIS "blind" pass and will make revisions, as needed.

Finally, Pass 3 will be made on the EPA IBM Mainframe during which a final revision of RF3 will be made including the following:

- Assignment of permanent reach numbers to all features,
- Construction of a "reach number change" transaction dataset for use by users who have already tied other water data to the current reach numbers,
- Rebuild the master RF3 dataset using the new data structures and new linework,
- Build a set of DLG transaction datasets which will give DLG the RF3 reach numbers, junctions, connectivity and other data, and
- Revise the mainframe software that currently uses RF3.

The final version of RF3 will meet the requirements of EPA and other Federal Agencies for a national hydrographic spatial database. It will be totally compatible with the new USGS Digital Line Graph Enhanced data. The RF3 data will be available in formats for IBM mainframes, GIS workstations, and personal computers. As the first national hydrographic dataset which supports modeling and unique feature identifiers, RF3 is ideally positioned to play an important role in the National Spatial Data Infrastructure Framework dataset.