National Hydrography Dataset-Reach Addressing Database Design and Development Support

Logical Design Document - Version 2.1

Prepared for:

Thomas O. Dabolt
Environmental Protection Agency
OW/OWOW/AWPD/MB

Prepared by:

INDUS Corporation
1951 Kidwell Drive
Eighth Floor
Vienna, Virginia 22182

Reach Address Database

Contract Number: 50CMAA900048
Task Order Number: 56CMAA000006
Task Number: 001

March 2, 2005

Table of Contents

List of Figures

March 02, 2005

National Hydrography Dataset
Reach Addressing Database Design and Development Support
Logical Design

1.0 INTRODUCTION

The National Hydrography Dataset (NHD) is a database that interconnects and uniquely identifies the millions of stream segments or reaches that comprise the Nation's surface water drainage system. It is based upon the USGS 1988 1:100,000-scale Digital Line Graph (DLG) hydrography dataset integrated with reach-related information from the USEPA Reach File Version 3.0-Alpha release (RF3-Alpha). The NHD provides a national framework for assigning reach addresses to water quality related entities, such as dischargers, drinking water supplies, streams effected by fish consumption advisories, wild and scenic rivers, Clean Water Act Section 305(b) and 303(d) waterbodies, Designated Uses, etc. Reach addresses establish the locations of these entities relative to one another within the NHD surface water drainage network in a manner similar to street addresses. The assignment of reach addresses is accomplished through a process known as reach indexing.

Once linked to the NHD surface water drainage network by their reach addresses, the upstream/downstream relationships of the water quality related entities and any associated information about them can be analyzed using Geographic Information Systems (GIS) software. A GIS provides a set of computerized analytical and mapping tools that operate on geographic datasets, such as the NHD, land use, soils and population, that are organized in separate layers. A GIS can also be used to combine NHD-based network analysis with these other GIS layers to help better understand and display their respective effects upon water quality.

Furthermore, since the NHD provides a nationally consistent framework for addressing and analysis, water quality information linked to reach addresses by one organization can be shared with other organizations and easily integrated into many different types of applications to the benefit of all.

This document describes the logical design for the NHD Reach Address Database (RAD) Version 2.1 for storing NHD geometry and feature attribute tables along with reach indexed information as event tables. Event tables are the intermediary linkage between the NHD and EPA programmatic data stored in other databases such as Total Maximum Daily Load (TMDL) Tracking System, National Assessment Database, the Water Quality Standards Database (WQSDB), and Envirofacts. The goal for the NHD RAD is to facilitate access by Web-based GIS applications. The technical architecture for development of the national NHD RAD is the Environmental Systems Research Institute's (ESRI) Spatial Database Engine (ArcSDE) and Oracle Spatial for spatial layer data, and Oracle for attribute and auxiliary data. All NHD and event spatial data will be stored as both ArcSDE Binary and Oracle Spatial geometries to best facilitate a wide range of applications accessing the data.

2.0 APPLICABLE AND REFERENCED DOCUMENTS

This section lists reference documents that apply to the RAD. Listed editions were valid at the time of publication. All documents are subject to revision, but these specific editions govern the logical design concepts described in this document. Other documents listed here may reference additional information or background.

National Hydrography Dataset-Reach Addressing Database Design and Development Support System Requirements Document, Prepared by INDUS Corporation for the U.S. Environmental Protection Agency Office of Water. April 07, 2000

National Hydrography Dataset Concepts and Content – DRAFT. U.S. Geological Survey National Mapping Division. February, 2000.

National Hydrography Dataset-Reach Addressing Database Design and Development Support Physical Design Document - Version 1.1, Prepared by INDUS Corporation for the U.S. Environmental Protection Agency Office of Water. August 10, 2001.

National Hydrography Dataset-Reach Addressing Database Design and Development Support Logical Design Document - Version 1.1, Prepared by INDUS Corporation for the U.S. Environmental Protection Agency Office of Water. March 20, 2001.

Reach Indexing Tool for the National Hydrography Dataset (NHD RIT) Preliminary Design Document, Prepared by Research Triangle Institute for the U.S. Environmental Protection Agency Office of Water. November, 1999.

National Hydrography Dataset NHDinARC Data Model Version 1.04c. U.S. Geological Survey. September, 28, 1999.

ESRI White Paper - Spatial Database Engine. Environmental Systems Research Institute, Inc. April 1998.

ESRI White Paper – Mapping for the Data Warehouse. Environmental Systems Research Institute, Inc. September 1997.

ArcSDE Administration Guide, ArcSDE 8. Environmental Systems Research Institute, Inc. 1999.

Oracle Documentation Library, Release 8.1.7. Oracle. 2000.
ARC/INFO User's Guide, Dynamic Segmentation, Modeling Linear Features. Environmental Systems Research Institute, Inc. March, 1992.

3.0 DYNAMIC SEGMENTATION

This section describes how dynamic segmentation is implemented within ArcInfo, ArcSDE, and Oracle Spatial. The implementation is very similar within ArcSDE and Oracle Spatial, and the basic concepts are the same for any of these spatial storage methods.

3.1 THE ARCINFO DYNAMIC SEGMENTATION DATA MODEL

The dynamic segmentation data model provides the ability to store, display, query and analyze information associated with linear features, such as NHD reaches, without modifying the underlying coordinates of those features. Dynamic segmentation uses a known feature and a position or *measure* on it to represent data associated with linear features. To accomplish this, routes, and a measurement system for those routes, are defined for the linear features. The route itself is a linear feature that can have associated attributes known as *events*. Dynamic segmentation is the ability to link events to linear features using this route-measure format, and store, display, query, and analyze event information without segmenting or breaking the linear feature. Therefore multiple sets of event information can be associated to any part of a route without modification of the base coordinate data for the linear features on which the route system was built.

Figure 1 shows the ArcInfo dynamic segmentation data model. The linear features, NHD reaches, are represented as arcs within ArcInfo. These arcs have attributes contained within the Arc Attribute Table (AAT). A route system has been built on these reaches using the NHD reach identifier, RCH_CODE, as the route key. The route system is the collection of routes defined on the NHD reaches and is comprised of two feature attribute tables: the section table (SEC) and the route attribute table (RAT). A *section* may represent all or part of an arc and each section is assigned a *from* and *to* position defining its beginning and ending point along the arc. Each section is also assigned a starting and ending measure defining its position within the route. A *route* is an ordered collection of sections and their measures. Each route has an attribute key, RCH_CODE, and it is through this key and the user-assigned measures that events are displayed along a route. It is important to note that the measurement system used, along with the route key and all event table information, are user-maintained items. The measurement system for the route must be the same as the measurement system in the event tables. The measures for the NHD linear reach data are in *percent of arc* where each arc starts at the downstream point at zero and ends at the upstream point at 100. The exception to this rule is in branched reaches where the main stem has the 0 - 100 measure assigned as usual, but the non-main stem arcs are collectively assigned measures from 100-200. For purposes of the RAD, events can be *point* - single locations along the route, *linear* - denoted by a range of measures describing the linear extent of the event along the route, *waterbody* - denoted by a whole waterbody reach, or custom shapes - point, linear, or waterbody features not indexed to the NHD.

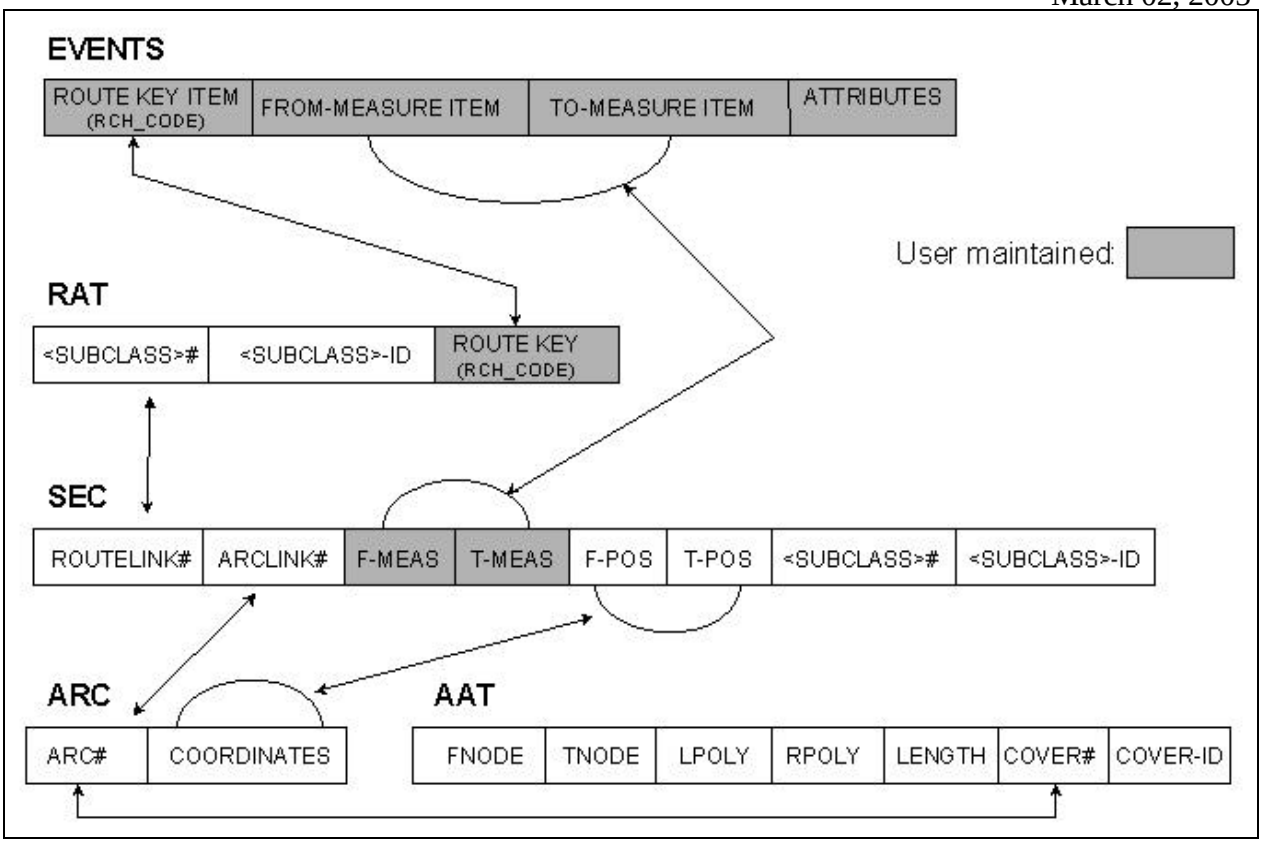


Figure 1: The ArcInfo Dynamic Segmentation Model

The event tables associated with NHD routes include an *Entity_ID* field. It is this field that is used to relate to other tables/databases containing information about the particular event, for example EPA programmatic databases.

The logical design of the NHD RAD does not implement the ArcInfo dynamic segmentation model within the database. Arcs, sections, polygons, and rxps (tables linking ArcInfo regions to polygons) are not loaded into Version 2.1 of the RAD. All point and route features (both NHD and Event) are loaded into the database as measured shapes and waterbody features are loaded in as area shapes. Both ArcSDE and Oracle Spatial implement similar Linear Referencing Systems that allow for the utilization of the measured shapes.

3.2 THE ARCSDE DYNAMIC SEGMENTATION DATA MODEL

The dynamic segmentation data model within ArcSDE is similar to that of ArcInfo in that it provides a means of measuring (or identifying locations) along a linear feature. The difference resides in how the measures are stored with the features. The ArcInfo dynamic segmentation model uses a route system with measures being stored on the sections relating the underlying arcs to the routes. In ArcSDE the measures are stored directly with the feature geometry (polylines) as a third dimension. The routes stored in the ArcInfo coverage format are converted to polylines upon load into the ArcSDE Oracle

database. Each record from the coverage route table (.RAT) is converted to a polyline record in ArcSDE. The records from the section table (.SEC) and arc table (.AAT) define the measures and geometry for the polylines stored in ArcSDE.

3.3 THE ORACLE SPATIAL DYNAMIC SEGMENTATION DATA MODEL

The dynamic segmentation data model within Oracle Spatial (referred to as a linear referencing system or LRS) is conceptually the same as within ArcSDE and ArcInfo. As with ArcSDE the measures in Oracle Spatial are stored directly with the feature geometry as a third dimension. The main difference between Oracle Spatial and ArcSDE is the Oracle table structures used to store the feature information. In Oracle Spatial the features are stored in the business table and in ArcSDE the features are stored in a separate Feature table linked by a common ID.

4.0 DATA MODELS

Data within the NHD RAD will be stored as both ArcSDE binary geometries and Oracle Spatial geometries. The storing of data in both formats allows for access to the NHD spatial data via a variety of applications. For some applications (mapping and interactive) access via ArcSDE is necessary, whereas for other applications (query based) access within the Oracle environment is necessary. Below are descriptions of how the NHD data model is implemented in both ArcSDE format and Oracle Spatial format.

4.1 THE ARCSDE DATA MODEL

ESRI's ArcSDE provides a continuous, non-tiled spatial model coupled to a relational database management system (RDBMS) such as Oracle. ArcSDE facilitates high-performance access to all data in the RDBMS, spatial and nonspatial. All features in an ArcSDE layer are indexed to optimize spatial search and retrieval. Indexes are created by dividing the layer into regular grid cells sized according to the geographic area that best represents, on average, the spatial area of interest such as a hydrologic cataloging unit (CU).

In ArcSDE, geometries or spatial data are stored as ArcSDE Large Object (LOB) data type. The ArcSDE data model consists of tables to maintain feature classes and datasets, spatial references, states, and versions. In addition to these tables, the ArcSDE data model supports the geodatabase system tables as well. The geodatabase model is not utilized by this implementation of the NHD RAD, and these tables will not be discussed here. Also, the NHD RAD is not implemented as a versioned database, and those tables related to ArcSDE versioning are also not discussed. For information regarding these tables, please refer to the ArcSDE Administration Guide, Appendix B.

The four ArcSDE tables used to maintain feature classes and datasets within the Oracle instance are described below. Two of these related tables are metadata tables:

The LAYERS table that maintains a list of all available layers in the database and includes the business table name, layer_id, column name for the *shape* column, layer description information, index grid size, data types and owner name.

The VERSION table that maintains the version of the ArcSDE server with which the Oracle database expects to operate.

ArcSDE spatially enables tables by adding a *shape* column. These spatially enabled tables are known as business tables. For each business table (layer) two other tables are created:

F<layer_ID> table	The feature table. It stores the shape geometry of the feature and related information such as area, length and type.
S<layer_ID> table	The spatial index table. It contains a record for each shape located within an index grid cell.

ArcSDE uses the integer values in the *shape* column as the key to relate, via a structured query language (SQL) join, to information in the F<layer_ID> table and the LAYERS metadata table. Therefore any table with a spatial column is a layer within ArcSDE. Business tables may or may not already exist within the Oracle database so any attribute information directly associated with each key may also be stored within the business table.

Figure 2 shows the relationships of these ArcSDE tables within the Oracle instance.

Each ArcSDE layer contains coordinate reference information that is maintained in the LAYERS table. The coordinate reference includes the coordinate system and the information needed to convert from real-world coordinates to the positive integer values stored by ArcSDE. ArcSDE assumes all data are in planar coordinates when calculating locations, areas or distances. Data must be scaled, and an offset calculated, to ensure all coordinates are loaded into ArcSDE as positive integers.

The source of the NHD is the USGS Feature Operational Database (FOD). The FOD is an Oracle/SDE database that maintains the NHD. These data are disseminated for use in "cans", ArcInfo coverage workspaces, which contain the data for all features that intersect with or are wholly within each hydrological cataloging unit (CU). The load process for the NHD RAD will consist of appending the NHD "cans" into the ArcSDE structure using a driver ArcInfo AML that makes calls to UNIX system scripts. Fields within layers that require ArcInfo topological relationships to populate are populated prior to load into the Oracle database. These processes allow for the loading of the feature route tables only and not loading the section and arc tables.

The layerimport command in the ARCPLOT extension of ARCSDE can be used to convert a route

system on an ArcInfo coverage to a measured line. A measured line is a set of one or more linear features called parts. Each linear feature is represented by two or more (x,y,m) points

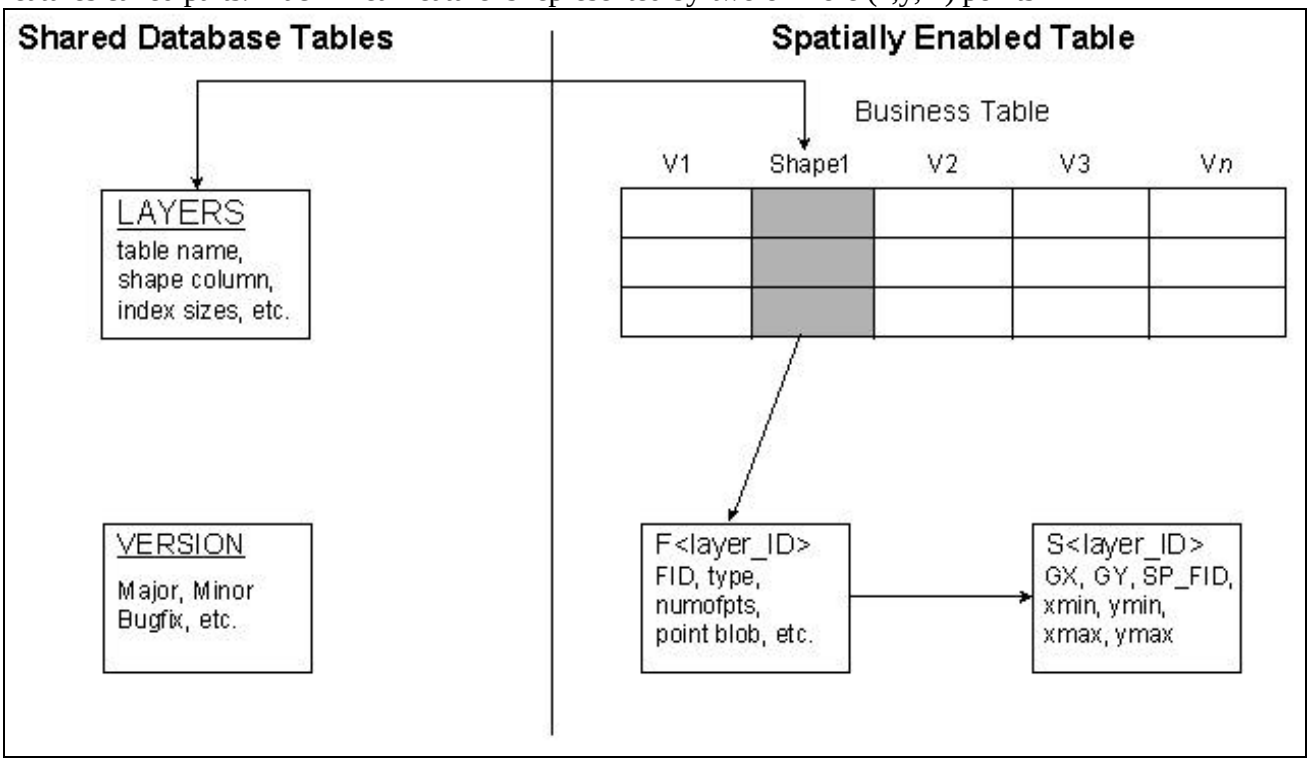


Figure 2: The Relationship of Tables in SDE

where the "x,y" values represent a location and the "m" value, a measure, defines a discrete location along the linear feature.

The items of interest from the section table include the F-MEAS and T-MEAS, which are used to calculate the "m" or measure values for event themes, and the F-POS and T-POS of the arc that hold the x,y location of the arc's nodes. The route key field from the route attribute table is one of the attributes in the measured line feature table.

4.2 THE ORACLE SPATIAL DATA MODEL

The Oracle Spatial data model is a hierarchical structure consisting of elements, geometries, and layers, which correspond to representations of spatial data. Layers are composed of geometries, which in turn are made up of elements. Oracle Spatial provides a SQL schema and functions that facilitate the storage, retrieval, update, and query of collections of spatial features in an Oracle database. The Oracle Spatial data model consists of tables to store spatial features (geometries) of a layer, a metadata view to store all metadata information, a spatial indexing mechanism, a set of operators and functions for performing area-of-interest and spatial join queries, and administrative utilities.

All features in an Oracle Spatial layer are spatially indexed to optimize spatial search and retrieval. A

spatial index is considered a logical index. The entries in the spatial index are dependent on the location of the geometries in a coordinate space. Once data have been loaded into the spatial tables, a spatial index must be created on the tables for efficient access to the data. In the RAD an R-tree index is built on each of the Oracle Spatial layers.

In Oracle Spatial, geometries are represented by either object-relational model or relational model. For the purpose of NHD data loading, the object-relational mechanism/model will be used because of its better performance over relational model and ability to support additional geometry types, such as arcs, circles, compound polygons, compound line strings, and optimized rectangles. The object-relational model uses a table with a single column of type MDSYS.SDO_GEOMETRY for the shape (or spatial) column and a single row per geometry instance. A geometry object is the representation of a spatial feature, modeled as an ordered set of primitive elements.

The USER_SDO_GEOM_METADATA view maintains a list of all Oracle Spatial layer names, column name for the shape column, coordinate system information (SRID), and DIMINFO. The value of SRID is available in MDSYS.CS_SRS table. If SRID is null, no coordinate system is associated with the geometry.

With coordinate system support in Oracle Spatial, one coordinate system can be converted to another coordinate system, and spatial functions, operators, and utilities provide correct and unambiguous results in whatever coordinate system the data are stored, particularly relating to measurements on the Earth's surface.

The ArcSDE engine can be used to load spatial data as Oracle Spatial geometry data type. The Layerimport command within the ARCPLOT extension can be used to convert a route system on an ArcInfo coverage to a measured line. A measured line is a set of one or more linear features called parts. Each linear feature is represented by two or more (x,y,m) points where the "x,y" values represent a location and the "m" value, a measure, defines a discrete location along the linear feature.
The items of interest from the section table include the F-MEAS and T-MEAS, which are used to calculate the "m" or measure values for event themes, and the F-POS and T-POS of the arc that hold the x,y location of the arc's nodes. The route key field from the route attribute table is one of the attributes in the measured line feature table.

5.0 NHD RAD INFORMATION CONTENT AND STRUCTURE

The NHD RAD will hold both the NHD and event table data for the EPA program system databases. These event tables will be related to the route systems in the NHD section of the database, with the exception of certain program systems that will contain non-NHD standard event layers. The data content for each section is discussed below.

5.1 NHD

The logical design for the NHD RAD is based on the NHDinARC format with the spatial elements adjusted to conform to the ArcSDE data model (see Figure 2). The implementation of the NHDinARC in the NHD RAD is called the NHDinSDE (see Appendix B). All spatial information will be unprojected (geographic projection) to facilitate application access by minimizing projection conversion issues.

The NHD will be represented as 10 ArcSDE business tables, 10 Oracle Spatial business tables (plus ancillary Oracle Spatial tables) and 6 Oracle tables. The business or spatially enabled tables and Oracle Spatial tables will consist of 4 measured shapes, LINEARRCH, DRAIN, DRAIN_RCH (containing reached drains), and LINEARLM; 4 polygon shapes, WB, AREARCH, AREALM, and DUUDOM; and 2 point shapes, NODES and POINTS. The 5 Oracle tables will consist of FREL, RCL, DUU2FEA, FCODE, DFLOW, and RFLOW.

There are two NHD metadata file types: CU, and quad. Each CU and quad may have multiple metadata files associated with them based on data revisions. The CU metadata files are loaded into the RAD and contained as CLOB data types. The quad level metadata are not loaded into the RAD and is available with the source NHD workspaces.

5.2 Event Data

Event table information will be aggregated programmatically such that there are three event layers; point, linear, and area corresponding to each program (e.g., 305(b), 303(d), and WQS). Each event layer will consist of shape data for each event. All three tables will consist of a common set of fields relating to the event and the reach indexed to, with the point event table adding P_MEAS and EOFFSET fields, and the linear event table adding F_MEAS, T_MEAS, and EOFFSET fields. For the case of events not indexed to the NHD there will be non-NHD standard event layers based on event type. These event layers will not contain information regarding reach reference.

Events are generally managed by a state organization or organization affiliated with a state, although there are exceptions. Both the organization and the geographic state will be stored for each event. In most cases these will be the same, but they may differ in cases such that tribes, state cooperators, and other non-state specific organizations manage the data. This distinction allows for both state based and management-based analysis.

In order to facilitate time series analysis within a program or across programs an additional set of layers (referred to as archive layers) contains both the current set of event data and older data that has been archived (replaced). These layers contain dates to indicate the time frame for when an event is considered valid, and when appropriate a date indicating the cycle for the data.

Event level metadata will be included as Oracle tables in the NHD RAD. Each program system's event

layer type (line, point, or area) will have a corresponding event metadata table. The archive layers will have an additional set of metadata tables by type as well. See the NHD RAD Physical Design Document for details regarding the table structures of the event layers and metadata tables.

6.0 CONCLUSION

The NHD RAD Physical Database Design Document will expand on this logical design and describe the database structure and design in detail. Although the database design addresses event maintenance, the database does not address NHD maintenance, which must be considered for later versions of the production database.

March 02, 2005

Appendix A

Definitions & Acronyms

AAT	Arc Attribute Table
ARC	The ArcInfo table that contains the feature geometry.
ArcSDE	A gateway to a multi-user commercial RDBMS , for example Oracle, Microsoft SQL Server, Informix, and DB2. ArcSDE is an open, high-performance spatial data server that employs client/server architecture to perform efficient spatial operations and manage large, shared geographic data.
BEACH	Beaches Program
CU	Hydrologic Catalog Unit
DLG	Digital Line Graph
Designated Uses	Designated uses are the uses that water should support, such as drinking water supply, primary contact recreation (swimming), and aquatic life support. Each designated use has a unique set of water quality requirements or criteria that must be met for the use to be realized. States, Tribes, and other jurisdictions may designate an individual waterbody for multiple designated uses.
DWI	Drinking Water Intakes
Envirofacts	A national information system that provides a single point of access to data extracted from seven major EPA databases.
EOFFSET	Event offset in any event table. Used to reduce the number of events overlapping when the events are mapped by setting the offsets of events to different values.
Event	The extent of the surface waters (reaches) encompassed by a single identifier in a program, for example, 303(d) listed waters.
EPA	Environmental Protection Agency
ESRI	Environmental Systems Research Institute
F_MEAS	From measure in a linear event table
FISH	Fish Consumption Advisories
FOD	Feature Operational Database
GIS	Geographic Information Systems
Geodatabase	An object oriented database that is hosted inside an RDBMS. Object behavior is implemented using validation rules, relationships, and topological associations.
GRTS	Grant Resource Tracking System
NDZ	No Discharge Zones

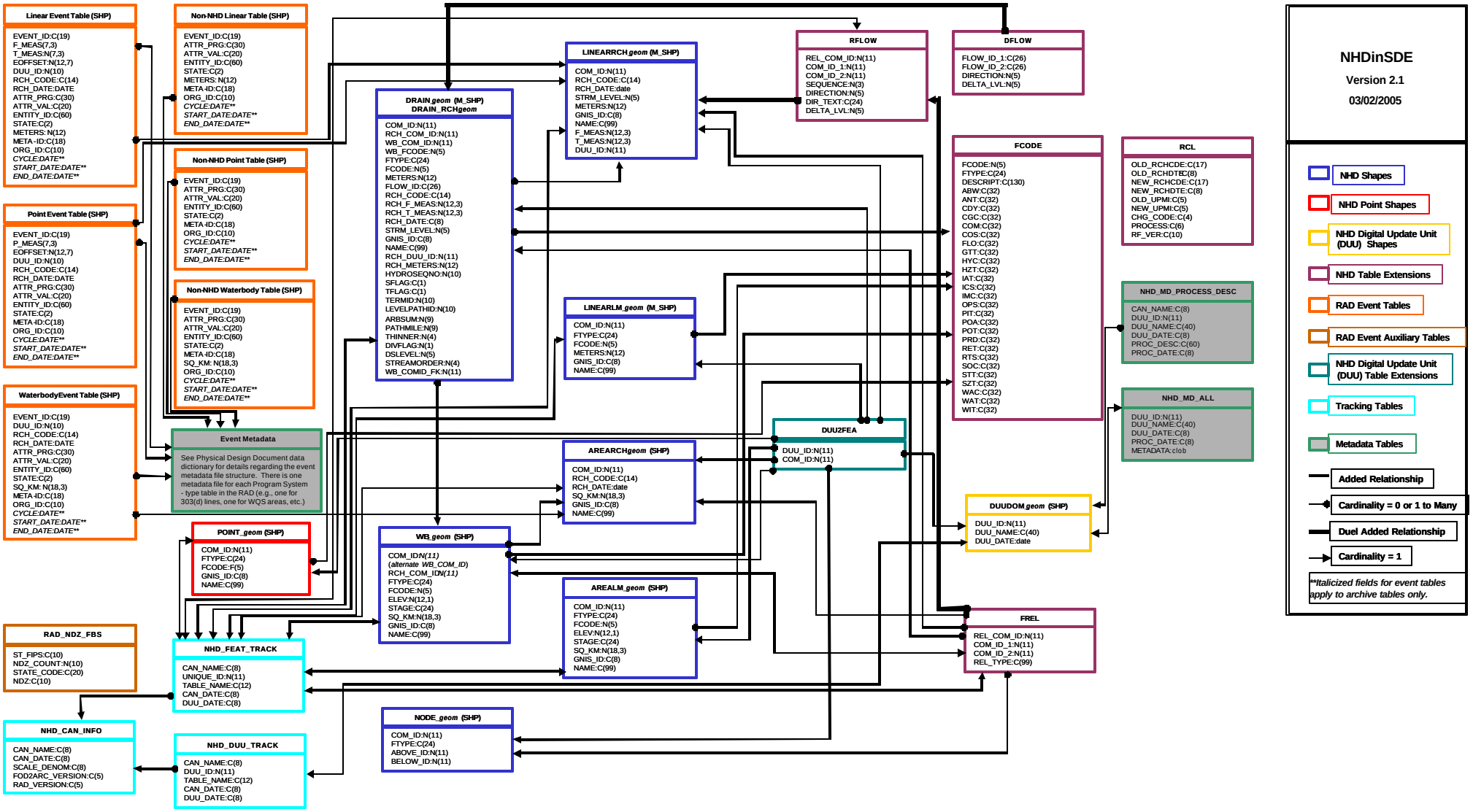
NHD	National Hydrography Dataset
NHDinARC	ArcInfo Data Model for NHD
NHDinSDE	ArcSDE Data Model for NHD
NHD RAD	NHD Reach Address Database
NHD RIT	NHD Reach Indexing Tool
OW	Office of Water
OWOW	Office of Wetlands, Oceans, and Watersheds
P_MEAS	Point measure in a point event table
PCS	Permit Compliance System
RAD	Reach Address Database
RAT	Route Attribute Table
RCH_CODE	A numeric code that uniquely identifies a reach. This 14-digit code has 2 parts: the first 8 digits are the hydrologic unit code for the cataloging unit in which the reach is located; the last 6 digits are a sequentially ordered, arbitrarily assigned number.
RDBMS	Relational Database Management System
RTP	Research Triangle Park
SDE	Spatial Database Engine
SEC	Section Table
SQL	Structured Query Language
T_MEAS	To measure in a linear event table
TMDL	Total Maximum Daily Load
USGS	United States Geological Survey
Versioning	A version is an alternative representation of the geodatabase that has an owner, a description, and a level of access (private, protected, and public).
WQSDB	Water Quality Standards Database
303(d)	Section 303(d) of the Clean Water Act requires states to identify waters that do not meet state water quality standards and develop Total Maximum Daily Load (TMDL) programs for these waters.
305(b)	Section 305(b) of the Clean Water Act requires that States and other jurisdictions submit a water quality report to EPA every two years. This report contains information about water quality conditions in rivers, lakes, estuaries,

wetlands, coastal waters, the Great Lakes, and ground water.

March 02, 2005

Appendix B

NHD RAD Object Diagram



NHD RAD Object Diagram