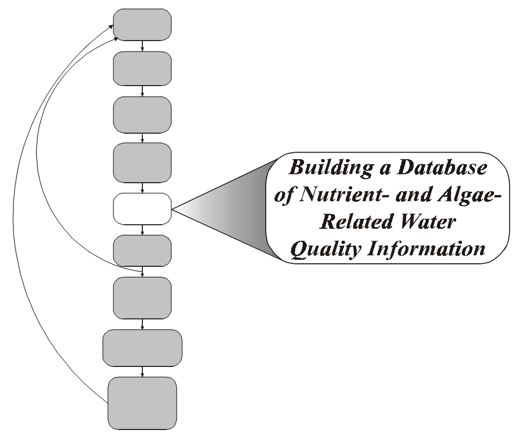
Chapter 5. Building a Database of Nutrient- and Algae-Related Water Quality Information



5.1 INTRODUCTION

A database of relevant water quality information can be an invaluable tool to States and Tribes as they develop nutrient criteria. Existing data may provide considerable information that is specific to the region where criteria are to be set. First the data must be located, then the suitability of the data (type and quality) ascertained before they are to be used for historical reconstruction of water quality parameters. It is also important to determine how the data were collected to make future monitoring efforts compatible with earlier approaches.

Databases operate much like spreadsheet applications, but have greater capabilities. While spreadsheets analyze and graphically display small quantities of data, databases store and manage large quantities of data and allow viewing and exporting of data sorted in a variety of ways. Databases can be used to organize existing information, store newly gathered monitoring data, and manipulate data as criteria are being developed. Databases can sort data for export into statistical analyses programs, spreadsheets, and graphics programs. This chapter will discuss the role of databases in nutrient criteria development, and provide a brief review of existing sources of nutrient-related water quality information for streams and rivers.

5.2 DATABASES AND DATABASE MANAGEMENT

This section describes general database structure and provides detailed information on relational and GIS databases. A database is a collection of information related to a particular subject or purpose. Databases are arranged so that they divide data into separate electronic repositories called tables. Data in tables can be viewed and edited, and new data can be added. A single datum is stored in only one table, but can be viewed from multiple locations. Updating one view of a datum will update it in all the various viewable forms. Each table should contain a specific type of information. Data from different tables can be viewed simultaneously according to the user-defined table relationships. That is, the relationship among data in different tables can be defined so that more than one table can be queried or reported, and accessed in a single view. Data stored in tables can be located and retrieved using queries. A query allows the user to find and retrieve only the data that meets user-specified conditions. Queries can also be used to update or delete multiple records simultaneously and to perform built-in or custom calculations of data. Data in tables can be analyzed and printed in specific layouts using reports. Data can be analyzed or presented in a specific way in print by creating a report.

To facilitate data manipulation and calculations, it is highly recommended that historical and present-day data be transferred to a relational database. A relational database is a collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. Each table (which is sometimes called a relation) contains one or more data categories in columns. Each row contains a unique instance of data for the categories defined by the columns. The organization of data into relational tables is known as the logical view of the database. In other words, the logical view is the form in which a relational database presents data to the user and the programmer (www.whatis.com/relation.htm). Relational databases are powerful tools for data manipulation and initial data reduction. They allow selection of data by specific, multiple criteria, and definition and redefinition of linkages among data components.

GISs are geo-referenced relational databases that have a geographical component (i.e., spatial platform) in the user interface. Spatial platforms associated with a database allow geographical display of sets of sorted data. Databases with spatial platforms are becoming more common. The system is based on premises that "pictures are worth thousands of words" and most data can be related to a map or other easily understood graphic. GIS platforms such as ArcView™, ArcInfo™, and MapInfo™ are frequently used to integrate spatial data with monitoring data for watershed analysis.

National Nutrients Database

The Nutrient Criteria Program has initiated development of a national relational database application that will be used to store and analyze nutrient data. The ultimate use of these data will be to derive ecoregion- and waterbody-specific numeric nutrient criteria ranges. Initially, EPA is developing a Microsoft Access™ application which will ultimately be populated with STORET Legacy Data, USGS NAWQA, NASQAN and Benchmark data, and other relevant nutrient data from universities, States/Tribes, and additional data rich entities. EPA is also developing a compatible, interactive system in an Oracle™ environment which allows for easy web-accessibility, geo-referencing/GIS compatibility, and data analysis on both a State/Tribal, regional, and national basis. The total amount of existing nutrient data nationally is large (>20 gigabytes), and it is anticipated that more data will be entered into the system. The Oracle™ application can easily manage large quantities of data and will provide ample room for expansion as more data are collected. Both the Access™ and the Oracle™ database applications are being designed for compatibility with EPA's latest edition of STORET to avoid duplication of effort for users of STORET and the Nutrients database application. Considerable efforts are also being made to assure compatibility with other database systems (e.g., WQS and RAD) currently being developed in EPA's Office of Water. The Microsoft Access™ application will be available in January 2000; the Oracle™ application will be online in the spring of 2000.

5.3 COLLECTING EXISTING DATA

In some States/Tribes, historical data on streams and rivers are already available. These data can be used to identify reference streams and begin development of potential nutrient criteria. Data should be compiled in a format that is easily imported into database and spreadsheets. Ideally, data will be compiled in the Nutrients Database described above. Potential data sources for river and stream nutrient data that will be useful for developing criteria are discussed below. These data sources contain extensive water quality data, however, data collection should not be limited to these sources. Collection of scientifically sound water quality data from any reliable source is encouraged.

POTENTIAL DATA SOURCES

Potential sources of data include water quality monitoring data from Federal, State, Tribal and local water quality agencies; university studies; and volunteer monitoring information. The data sources described in this section do not encompass the full extent of available data sources. Many State/Tribal, and Federal programs that are regional or site-specific are excellent data sources, but are not included in this discussion.

EPA Water Quality Data

EPA has many programs of national scope that focus on collection and analysis of water quality data. The following presents information on several of the databases and national programs that may be useful to water quality managers as they compile data for criteria development.

STORET

STOrage and RETrieval system (STORET) is EPA's national database for water quality and biological data. EPA's original STORET System, operated continuously since the 1960s, was historically the largest repository of water quality data in the nation. This legacy mainframe-based system will cease to exist in the year 2000. In its place, EPA will support two independent, web-accessible databases. The older database, called the STORET Legacy Data Center (LDC) is the repository of all data held in EPA's original STORET System as of the end of 1998. The newer, modernized database, simply called STORET, is designed as a replacement for the original STORET System. It is the repository for more current data, and offers major improvements in database content and quality control documentation.

Interested parties may view both databases on the World-Wide-Web, where the capability will exist to produce printed reports and download data files. Queries for data via the web will be designed for use by the general public and will require no special training or software. The web site will be announced in the first quarter of FY2000.

STORET (the new STORET system) is a compendium of data supplied by Federal, State, and local organizations which evaluates environmental conditions in the field. The data in STORET is organized by both geographic location and data ownership. Every field study site is identified by at least one latitude/longitude and, where appropriate, also by State/Province, County, drainage basin, and stream reach. Monitoring activities recorded include field measurements, habitat assessments, water and sediment samples, and biological population surveys. Records cover the complete spectrum of physical properties, concentrations of substances, and abundance and distribution of species observed during biological monitoring. STORET is designed for maximum compatibility with commercial software, including Geographic Information Systems such as the ESRI ArcView package, and statistical packages such as PC SAS. STORET download files import easily into all standard spread sheet packages.

Further information about STORET may be obtained by e-mailing STORET@epa.gov, or telephoning toll-free at 1-800-424-9067.

National Surface Water Survey (NSWS)
EPA's National Surface Water Survey consists of two parts: the National Lake Survey and the National Stream Survey. The purpose of the National Lake Survey is to quantify, with known statistical confidence, the current status, extent, and chemical and biological characteristics of lakes in regions of the United States that are potentially sensitive to acidic deposition. The purpose of the National Stream Survey (NSS) is to determine the percentage, extent, and location of streams in the United States that are presently acidic or have low-acid neutralizing capacity and may, therefore, be susceptible to future acidification, as well as to identify streams that represent important classes in each region for possible use in more intensive studies or long-term monitoring. The NSS provides an overview of stream water quality chemical characterisitics in regions of the United States that are expected, on the basis of previous alkalinity data, to contain predominantly low-acid neutralizing capacity waters (EPA website http://www.epa.gov/ceisweb1/ceishome/ceisdocs/usguide/prog(56).htm).

Environmental Monitoring and Assessment Program (EMAP)
The Environmental Monitoring and Assessment Program is an EPA research program designed to develop the tools necessary to monitor and assess the status and trends of national ecological resources (see EMAP Research Strategy on the EMAP website: www.epa.gov/emap). EMAP's goal is to develop the scientific understanding for translating environmental monitoring data from multiple spatial and temporal scales into assessments of ecological condition and forecasts of future risks to the sustainability of the Nation's natural resources. EMAP's research supports the National Environmental Monitoring Initiative of the Committee on Environment and Natural Resources (CENR) (www.epa.gov/emap/). Data from the EMAP program can be downloaded directly from the EMAP website (www.epa.gov/emap/). The EMAP Data Directory contains information on available data sets including data and metadata (language that describes the nature and content of data). Current status of the data directory as well as composite data and metadata files are available on this website.

Clean Lakes Program (CLP)
The EPA Clean Lakes Program was initiated to assess water quality in impaired public lakes and reservoirs and to restore these systems where appropriate. CLP included a monitoring and assessment component to identify the efforts needed to restore water quality. Lakes in this program were selected because they were perceived to have water quality impairment. Major tributaries into lakes and reservoirs included in this program were sampled on a regular basis. EPA encouraged States in its May 1996 section 319 nonpoint source guidance to use section 319 funds to fund eligible activities that might have been funded in previous years in the CLP under Section 314. Data from this program may be useful for positioning river and stream systems on a nutrient gradient continuum, but are unlikely to provide data for reference stream reaches. Information about EPA's CLP can be found at the website: http://www.epa.gov/owowwtr1/lakes/lakes.old.html.

Ecological Data Application System (EDAS)
EDAS is EPA's program-specific counterpart to STORET. EDAS was developed by EPA's Office of Water to manipulate data obtained from biological monitoring and assessment and to assist States/Tribes in developing biocriteria. It contains built-in data reduction and recalculation queries that are used in

biological assessment. The EDAS database is designed to enable the user to easily manage, aggregate, integrate, and analyze data to make informed decisions regarding the condition of a water resource. Biological assessment and monitoring programs require aggregation of raw biological data (lists and enumeration of taxa in a sample) into informative indicators. EDAS is designed to facilitate data analysis, particularly the calculation of biological metrics and indexes. Pre-designed queries that calculate a wide selection of biological metrics are included with EDAS. Future versions of EDAS will include the capability to upload data to, and download data from, the distributed version of modernized STORET. EDAS is not a final data warehouse, but is a program or project-specific customized data application for manipulating and processing data to meet user requirements. The EDAS application is currently under development; more information will be available at a later date through the EPA website.

USGS (U.S. Geological Survey) Water Data

The USGS has national and distributed databases on water quantity and quality for waterbodies across the nation. Much of the data for rivers and streams are available through the National Water Information System (NWIS). These data are organized by state, Hydrologic Unit Codes (HUCs), latitude and longitude, and other descriptive attributes. Most water quality chemical analyses are associated with an instantaneous streamflow at the time of sampling and can be linked to continuous streamflow to compute constituent loads or yields. The most convenient method of accessing the local data bases is through the USGS State representative. Every State office can be reached through the USGS home page on the Internet at URL http://www.usgs.gov/wrd002.html.

HBN and NASQAN

USGS data from several national water quality programs covering large regions offer highly controlled and consistently collected data that may be particularly useful for nutrient criteria analysis. Two programs, the Hydrologic Benchmark Network (HBN) and the National Stream Quality Accounting Network (NASQAN) include routine monitoring of rivers and streams during the past 30 years. The HBN consisted of 63 relatively small, minimally disturbed watersheds. HBN data were collected to investigate naturally-induced changes in streamflow and water quality and the effects of airborne substances on water quality. The NASQAN program consists of 618 larger, more culturally influenced watersheds. NASQAN data provides information for tracking water-quality conditions in major U.S. rivers and streams. The watersheds in both networks include a diverse set of climatic, physiographic, and cultural characteristics. Data from the networks have been used to describe geographic variations in water-quality concentrations, quantify water-quality trends, estimate rates of chemical flux from watersheds, and investigate relations of water quality to the natural environment and anthropogenic contaminant sources. Since 1995, the NASQAN Program has focused on monitoring the water quality of four of the Nation's largest river systems—the Mississippi (including the Missouri and Ohio), the Columbia, the Colorado, and the Rio Grande. NASQAN currently operates a network of 40 stations in which the concentration of a broad range of chemicals—including pesticides and trace elements—and stream discharge are measured.

Alexander and others (1996) assembled much of the historical water-quality and streamflow data collected by the NASQAN and HBN on two CD-ROMs, including supporting documentation and quality assurance information (see Internet URL http://wwwrvares.er.usgs.gov/wqn96/). These data are collectively referred to as Water-Quality Networks (WQN). The CD-ROMs are designed to allow users to efficiently browse text files and retrieve data for subsequent use in user-supplied software including spreadsheet, statistical analysis, or geographic information systems. The data may be extracted from one of the CD-ROMs (the "DOS disc") using the supplied DOS-based software, and output in a variety of formats. This software allows the user to search, retrieve, and output data according to user-specified requirements. Alternatively, the ASCII form of the WQN data may be accessed on a second CD-ROM (the "ASCII disc") from user-supplied software including a Web browser, spreadsheet, or word processor.

A comprehensive review of sources, concentrations, and loads of nutrients in the Mississippi River Basin was completed by USGS under the Committee of Environmental Natural Resources. The review focused on analyzing issues related to the Gulf of Mexico hypoxia. Much of Topic 3, Flux and sources of nutrients in the Mississippi-Atchafalaya River Basin, includes data and analysis that could be useful for the development of nutrient criteria in large river systems, such as the Mississippi River. Results of this effort, which was led by the National Oceanic and Atmospheric Administration, have been published and can also be found at the Internet site http://www.nos.noaa.gov/products/pubs_hypox.html.

NAWQA

The USGS National Water-Quality Assessment (NAWQA) Program is building a third national database of stream quality information from data collected and analyzed for more than 50 river basins and aquifer systems, called Study Units, across the Nation. NAWQA studies are based on a complex sampling design that targets specific land uses and hydrologic conditions in addition to assessing the most important aquifers and large streams and rivers in each area studied. Gilliom and others (1995) describe the NAWQA sampling design in detail. A comprehensive data screening, computer retrieval, and review of existing data on nutrients in streams was completed for each of the first 20 Study Units (Mueller et al. 1995). A major component of the sampling design for streams is to target specific watersheds influenced primarily by a single dominant land use (agricultural or urban) that is important in a particular area of the United States. Some of the watersheds were selected as undeveloped areas relative to the rest of the Study Unit to use in comparative analysis of land-use effects on water quality. Water-quality data collection during 1992-1996 include analyses of eight nutrient species from about 8,500 samples of streams and rivers in the first 20 Study Units. A data set used for national synthesis of water quality has been compiled and can be viewed and downloaded via the Internet URL http://wwwrvares.er.usgs.gov/nawqa/nutrient.html. Mueller and others (1997) describe quality control of the NAWQA stream data and Mueller (1998) provides a rigorous assessment of the quality of these data.

WEBB

The Water, Energy, and Biogeochemical Budgets (WEBB) program was developed by USGS to study water, energy, and biogeochemical processes in a variety of climatic/regional scenarios. Five ecologically diverse watersheds, each with an established data history, were chosen. This program may prove to be a rich data source for ecoregions in which the five watersheds are located. Many publications on the WEBB project are available. See the USGS website for more details (http://water.usgs.gov/nrp/webb/about.html).

USDA

Agricultural Research Service (ARS)

ARS houses Natural Resources and Sustainable Agricultural Systems, which has seven national programs to examine the effect of agriculture on the environment. The program on Water Quality and

Management addresses the role of agriculture in nonpoint source pollution through research on Agricultural Watershed Management and Landscape Features, Irrigation and Drainage Management Systems, and Water Quality Protection and Management Systems. Research is conducted across the country and several models and databases have been developed. Information on research and program contacts is listed on the website (http://www.nps.ars.usda.gov/programs/nrsas.htm).

Forest Service

The Forest Service has designated research sites across the country, many of which are Long Term Ecological Research (LTER) sites. Many of the data from these experiments are available in the USFS databases located on the website (http://www.fs.fed.us/research/). Most of the data are forest-related, but may be of use for determining land uses and questions on silviculture runoff.

National Science Foundation (NSF)

The National Science Foundation funds projects for the LTER Network. The Network is a collaboration of over 1,100 researchers investigating a wide range of ecological topics at 24 different sites nationwide. The LTER research programs are not only an extremely rich data source, but also a source of data available to anyone through the Network Information System (NIS), the NSF data source for LTER sites. Data sets from sites are highly comparable due to standardization of methods and equipment. Data can be accessed from the website http://www.lternet.edu/research/data/nis/.

U.S. Army Corps of Engineers (COE)

The U.S. Army Corps of Engineers is responsible for more than 750 reservoirs. Many have extensive monitoring data that could contribute to the development of nutrient criteria for tributaries to those reservoirs. The COE focuses more on water quantity issues than on water quality issues. As a result, much of the river and stream system data collected by the COE does not include nutrient or algal constituents. Nonetheless, the COE does have a large water sampling network and supports USGS and EPA monitoring efforts in many programs. A list of the water quality programs that the COE actively participates in was compiled in 1997. This information can be found at the website: http://cw71.cw-wc.usace.army.mil/wqinfo/wq98sem/ANNWQMGT.HTM.

U.S. Department of the Interior, Bureau of Reclamation (BuRec)

The Bureau of Reclamation manages many irrigation and water supply reservoirs in the West, some of which may have operational data available. These data focus on water supply information and limited water quality data. However, real time flow data are collected for rivers supplying water to BuRec, which may be useful for the flow component of criteria development. These data can be gathered on a site-specific basis from the BuRec website: www.usbr.gov. Extensive remote sensing data are available from the website: http://wais.rsgis.do.usbr.gov/html/rsgig_wq.html.

State/Tribal Monitoring Programs

Most states monitor some subset of stream and river systems within their borders for algal and nutrient variables. Data collected by State/Tribal water quality monitoring programs can be used for nutrient criteria development. These data should be available from the agencies responsible for monitoring.

Volunteer Monitoring Programs

Many States have volunteer water quality monitoring programs. Some programs are state-sponsored, while others are independent organizations such as Adopt-A-Stream. Citizens in many areas donate their time, money, or experience to aid State, Tribal, and local governments in collecting water quality data. Volunteers analyze water samples for DO (dissolved oxygen), nutrients, pH, temperature, and a host of other water constituents; evaluate the health of stream habitats and aquatic biological communities; note stream-side conditions and land uses that may affect water quality; catalogue and collect beach debris; and restore degraded habitats.

State and local agencies may use volunteer data to screen for water quality problems, establish trends in waters that would otherwise be unmonitored, and make planning decisions. Volunteers benefit from learning more about their local water resources and identifying what conditions or activities might contribute to pollution problems. As a result, volunteers frequently work with clubs, environmental groups, and State/Tribal or local governments to address problem areas.

The EPA supports volunteer monitoring and local involvement in protecting our water resources. EPA support takes many forms including: sponsoring national and regional conferences to encourage information exchange among volunteer groups, government agencies, businesses, and educators; publishing sampling methods manuals for volunteers; producing a nationwide directory of volunteer programs; and providing technical assistance (primarily on quality control and lab methods) and Regional coordination through the ten EPA Regional offices. In addition, grants to States/Tribes that can be used to support volunteer monitoring in lakes and for nonpoint source pollution control are managed by the EPA Regions (http://www.epa.gov/OWOW/monitoring/volunteer/epavm.html).

Adopt-A-Stream

The Adopt-A-Stream Foundation (AASF) is a non-profit organization that works to increase public awareness and involvement in water quality issues, stream enhancement, and environmental education. Their two main areas of focus are Environmental Education and Habitat Restoration. AASF seeks to protect streams through volunteer work, encouraging school and community groups, sports clubs, civic organization, and individuals to become “Streamkeepers.” “Adoption” of a stream requires that volunteers provide long-term care of the stream and establish stream monitoring, restoration, and community-wide environmental education activities. AASF provides education materials, classes, and tools for monitoring. Data collected through the volunteer monitoring associated with Adopt-A-Stream is usually site-specific, focusing on a single stream. However, if volunteers have been properly trained, the data collected may be useful in helping identify streams at risk for nutrient problems. The AASF website contains additional information on this organization and data they may be able to provide (http://www.streamkeeper.org/).

American Heritage Rivers

The American Heritage Rivers Initiative is a program launched by President Clinton to help communities restore their local waters and waterfront areas. Participation is voluntary and must be initiated by the community. To date, fourteen rivers have been designated on the basis of historical, economic, and environmental considerations. One goal of the program is to develop additional information that can be used by communities to improve any river system. Through the American Heritage Rivers website (http://www.epa.gov/OWOW/heritage/rivers.html), valuable information about our nation's rivers is easily available to everyone. Information organized geographically on flood events, population change, road networks, condition of water resources, and partnerships already at work in the area is available. Additionally, customized maps and environmental and educational assessment models will be made available through this initiative.

Electric Utilities
Many electric utilities own reservoirs for hydroelectric power generation, and are required to monitor the reservoirs' water quality. The largest of these, the Tennessee Valley Authority (TVA), has extensive chemical and biological monitoring data from most of its reservoirs from the early 1980s to the present. Data collected in conjunction with hydroelectric reservoirs must be gathered from the facility owners or managers.

Drinking Water Facilities
Many local drinking water facilities are supplied from river systems. These facilities continuously monitor some water quality parameters at the intake pipe. Nutrients are infrequently monitored by most of these facilities, but supplemental data, i.e., turbidity, pH, and flow are usually measured. These data may not provide the necessary parameters for deriving criteria, but may be very useful in combination with State/Tribal water quality monitoring data to develop criteria. Data from these facilities should be accessed locally for the waterbody of concern.

Academic and Literature Sources
Many research studies are conducted by academic institutions that may provide data useful for developing nutrient criteria. Much of the research conducted by the academic community concentrates on unimpaired or minimally impaired systems. While data collected from these sources may not be directly representative of the population of stream systems within an ecoregion, they could be useful for identifying reference conditions. Academic research also tends to be site-specific and span a limited number of years, although data for some systems may span 20 years or more. Academic research data should be available from researchers and the scientific literature.

QUALITY OF HISTORICAL DATA

The value of older historical data sets is a recurrent problem because data quality is often unknown. Knowledge of data quality is also problematic for long-term data repositories such as STORET and long-term State databases, where objectives, methods, and investigators may have changed many times over the years. The most reliable data tend to be those collected by a single agency, using the same protocol, for a limited number of years. Supporting documentation should be examined to determine the consistency of sampling and analysis protocols. Investigators must determine the acceptability of data contained in large, heterogeneous data repositories. Considerations and requirements for acceptance of these data are described below.

Location
STORET and USGS data are geo-referenced with latitude, longitude, and Reach File 3 (RF3) codes. Geo-reference data can be used to select specific locations, or specific USGS Hydrologic Units. In addition, STORET often contains a site description. Knowledge of the rationale and methods of site

selection from the original investigators may supply valuable information. Metadata of this type, when known, is frequently stored within large long-term databases.

Variables and Analytical Methods

Thousands of variables are recorded in database records. Each separate analytical method yields a unique variable. For example, five ways of measuring TP results in five unique variables. We do not recommend mixing analytical methods in data analyses because methods differ in accuracy, precision, and detection limits. Data analyses should concentrate on a single analytical method for each parameter of interest. Selection of a particular "best" method may result in too few observations, in which case it may be more fruitful to select the most frequently used analytical method in the database. Data may have been recorded using analytical methods under separate synonymous names, or analytical methods incorrectly entered when data were first added to the database. Review of recorded data and analytical methods recorded by knowledgeable personnel is necessary to correct these problems.

Laboratory Quality Control (QC)

Laboratory QC data (blanks, spikes, replicates, known standards, etc.) are infrequently reported in larger data repositories. Records of general laboratory quality control protocols and specific quality control procedures associated with specific datasets are valuable in evaluating data quality. However, premature elimination of lower quality data can be counterproductive, because the increase in variance caused by analytical laboratory error may be negligible compared to natural variability or sampling error, especially for nutrients and related water quality parameters. However, data of uncertain and undocumented quality should not be accepted.

Data Collecting Agencies

Selecting data from particular agencies with known, consistent sampling and analytical methods and known quality will reduce variability due to unknown quality problems. Requesting data review for quality assurance from the collecting agency will reduce uncertainty about data quality.

Time Period

Long-term records are critically important for establishing trends. Determining if trends exist in the time series database is also important for characterizing reference conditions for nutrient criteria. Length of time series data needed for analyzing nutrient data trends is discussed in Chapter 4.

Index Period

An index period for estimating average concentrations can be established if nutrient and water quality variables were measured through seasonal cycles. The index period may be the entire year or the summer growing season. The best index period is determined by considering stream characteristics for the region, the quality and quantity of data available, and estimates of temporal variability (if available). Additional information and considerations for establishing an index period are discussed in Chapter 4.

Representativeness

Data may have been collected for specific purposes. Data collected for toxicity analyses, effluent limit determinations, or other pollution problems may not be useful for developing nutrient criteria. Further, data collected for specific purposes may not be representative of the region or stream classes of interest. The investigator must determine if stream systems or a subset of the stream systems in the database are representative of the population of stream systems to be characterized. If a sufficient sample of representative systems cannot be found, then a new survey will be necessary (see Chapter 4).

Gather New Data
New data should be gathered following the sampling design protocols discussed in Chapter 4. New data collection activities for developing nutrient criteria should focus on filling in gaps the database and collecting regional monitoring data. Data gathered under new monitoring programs should be imported into databases or spreadsheets and merged with existing data for criteria development.

Data Reduction
Data reduction requires a clear idea of the analysis that will be performed and a clear definition of the sample unit for the analysis. For example, a sample unit might be defined as "a watershed during July-August". For each variable measured, a mean value would then be estimated for each watershed in each July-August index period on record. Analyses are then done with the observations (estimated means) for each sample unit, not with the raw data. Steps in reducing the data include:

- Selecting the long-term time period for analysis;
- Selecting an index period;
- Selecting relevant chemical species;
- Identifying the quality of analytical methods;
- Identifying the quality of the data recorded; and
- Estimating values for analysis (mean, median, minimum, maximum) based on the reduction selected.

QUALITY ASSURANCE/QUALITY CONTROL

The validity and usefulness of data depend on the care with which they were collected, analyzed and documented. The EPA provides guidance on data quality assurance (QA) and quality control (USEPA 1998b) to assure the quality of data. Factors that should be addressed in a QA/QC plan are elaborated below. The QA/QC plan should state specific goals for each factor and should describe the methods and protocols used to achieve the goals. The five factors discussed below are: representativeness, completeness, comparability, accuracy and precision.

Representativeness
Sampling program design (when, where, how you sample) should produce samples that are *representative* or typical of the environment being described. Sampling designs for developing nutrient criteria are addressed in Chapter 4.

Completeness
Data sets are often incomplete because of spilled samples, faulty equipment, and/or lost field notebooks. A QA/QC plan should describe how complete the data set must be in order to answer the questions posed (with a statistical test of given power and confidence) and the precautions being taken to ensure that completeness. Data collection procedures should document the extent to which these conditions have been met. Incomplete data sets may not invalidate the collected data, but may reduce the rigor of statistical analyses. Therefore, precautions should be taken to ensure data completeness. Precautions to

ensure completeness may include collecting extra samples, having back-up equipment in the field, installing alarms on freezers, copying field notebooks after each trip, and/or maintaining duplicate sets of data in two locations.

Comparability

In order to compare data collected under different sampling programs or by different agencies, sampling protocols and analytical methods must demonstrate comparable data. The most efficient way to produce comparable data is to use sampling designs and analytical methods that are widely used and accepted.

Accuracy

To assess the accuracy of field instruments and analytical equipment, a standard (a sample with a known value) must be analyzed and the measurement error or bias determined. Internal standards should periodically be checked with external standards provided by acknowledged sources. At Federal, State, Tribal, and local government levels, the National Institute of Standards and Technology (NIST) provides advisory and research services to all agencies by developing, producing, and distributing standard reference materials. For calibration sevices and standards see: http://ts.nist.gov/ts/htdocs/230/233/calibration/home.html.

Standards and methods of calibration are typically included with turbidity meters, pH meters DO meters, and DO testing kits. The USGS, the EPA and some private companies provide reference standards or QC samples for nutrients. Reference standards for chlorophyll are also available from the EPA and some private companies, although chlorophyll standards are time and temperature sensitive because they degrade over time.

Variability

Natural variability rather than imprecision in the method used, is usually the greatest source of error in the constituent measured. The variability in field measurements and analytical methods should be demonstrated and documented to identify the source of variability when possible. EPA QA/QC guidance provides an explanation and protocols for measuring sampling variability (USEPA 1998b). Methods for creating a chlorophyll standard to determine if the spectrophotometer is measuring chlorophyll consistently from one year to the next or from the beginning to the end of an analytical run are described in Wetzel and Likens (1991). In addition, a large number of replicates for each sample time and site must be collected because the largest source of variation is likely to be natural (i.e., in the samples).