Center for Urban Environmental Studies
Northeastern University, Boston, MA 02115

TECHNICAL REPORT NO. 5

DEVELOPMENT OF A RELATIONAL DATABASE FOR STUDYING ECOLOGICAL RESPONSE OF STREAMS TO ANTHROPOGENIC WATERSHED STRESSES AND STREAM MODIFICATIONS

Alena Bartošová, Ph.D.
Assistant Professional Scientist, Illinois State Water Survey

Ramanitharan Kandiah, Ph.D.
Post Doctoral Research Associate, Center for Urban Environmental Studies

Vladimir Novotny, Ph.D., P.E.
Primary Investigator

Project sponsored by the Grant No. R83-0885-010 to Northeastern University from the US EPA/NSF/USDA STAR Watershed Program

Bernice Smith
EPA Project Officer

Iris Goodman
EPA Program Director

Boston, MA
June 2005

Abstract

A relational database is developed to analyze relationships between the physical, chemical, and habitat parameters of a waterbody and its biological integrity. The database stores all the acquired physical, chemical, habitat and biological data for studied watersheds. Different protocols and scoring systems employed by various agencies and individual States represent the greatest challenge in bringing the data together in one database, especially biological and habitat data. A standardized data model enables convenient data manipulation. The structure of the relational database developed on a Microsoft SQL Server 2000 is described, and the data sources are summarized. The structure also enables direct calculation of numerous metrics from biological (taxonomic) data. The report includes sample queries from the database on Microsoft Access using ODBC. Issues related to the establishment of the databases are discussed.

Acknowledgements

The research contained in this report was sponsored by the U.S. Environmental Protection Agency/National Science Foundation/U.S. Department of Agriculture STAR Watershed Program by a Grant (No. R83-0885-010) to Northeastern University. The authors greatly appreciate this support. The findings and conclusions contained in this report are those of the authors and not of the funding agencies, nor the STAR program.

Table of Contents

List of Tables

List of Figures

Introduction

The project, "Ecological response of streams to anthropogenic watershed stresses and stream modifications" is aimed to study streams in various ecological regions throughout United States. Modeling an entire watershed that incorporates physical, chemical, habitat and biological aspects requires a significant quantity of data. In most cases, these data are not all available from the same source. While there are some extensive monitoring programs that have been established by federal agencies, other monitoring programs run by state agencies or in some cases by local citizens groups provide invaluable information. These separate monitoring programs often complement each other in terms of the locations, time period and parameters measured. A variety of data sources and formats led to developing a database that contains all acquired data both on reach and watershed level. A common structure enables the data to be used in a comparable way. The data will be used in evaluating integrity of the waterbody and assessing relative influence of individual factors on aquatic health. Karr *et al.* (1986) have documented that these factors can be grouped into five major classes: energy sources, chemical parameters, habitat parameter, flow regime, and biotic interactions. Altering any of these factors and parameters will have an impact on the whole ecosystem and the biota as its integral component. The database is designed to include data sampled at a waterbody site or reach from all of these classes, namely water and sediment quality data, habitat data, flow data, and biological data.

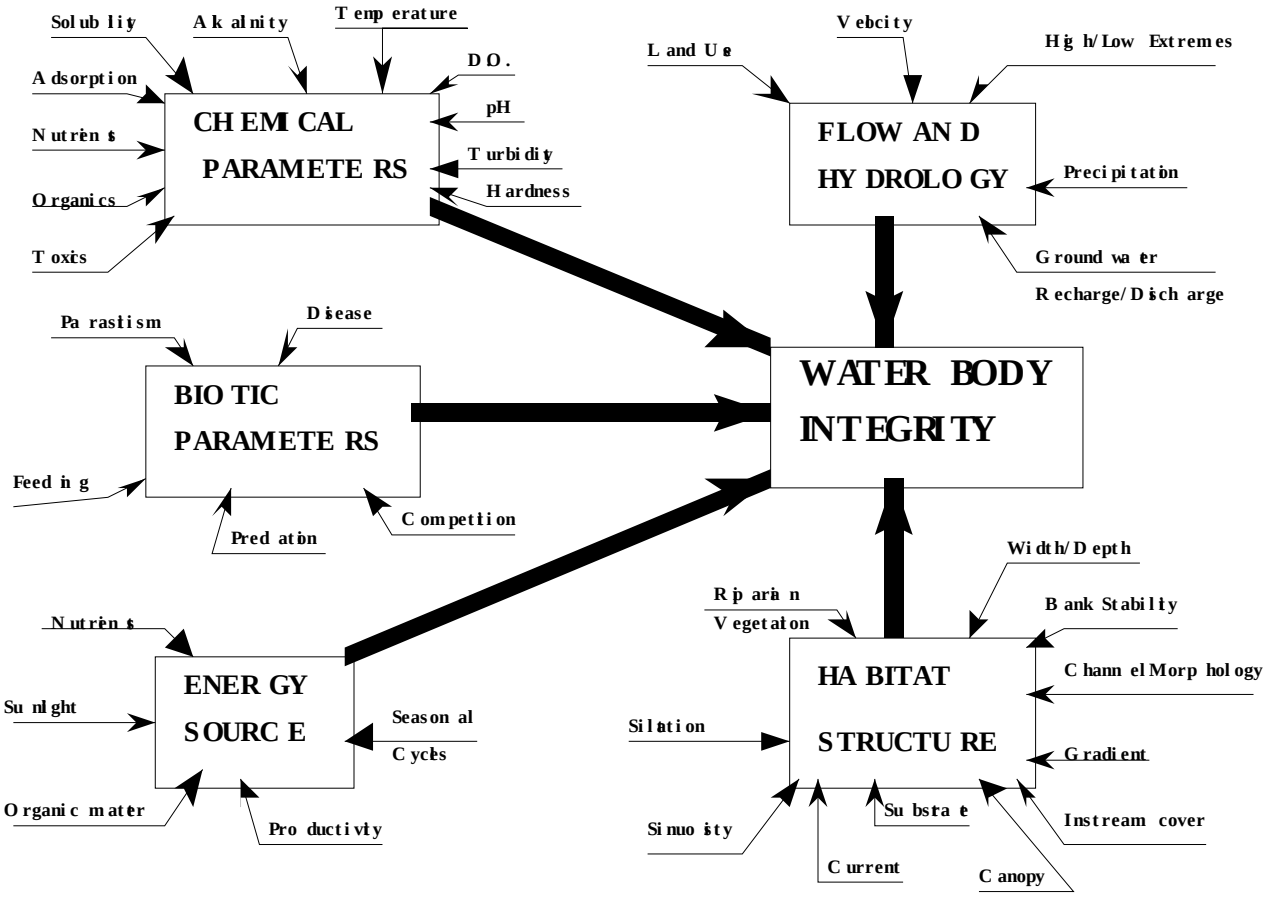

Figure 8 Factors that influence aquatic health (after Karr, *et al.*, 1986).

Additional environmental data important to the study are spatial in nature. The spatial data characterize an area contributing to a sampling site rather than the sampling site or a reach itself, e.g., land use distribution, soil types, or population density. GIS can be used to derive these characteristics based on location of sampling sites and data stored back in the database.

Database Structure

Environmental data needed for the study are acquired from different sources and consequently in different formats. Illinois State Water Survey (ISWS) faced a similar problem when compiling data for the Fox River Watershed Investigation (McConkey *et al.* 2004). Fox River (tributary to Illinois River) is one of the watersheds selected for this study. The relational database FoxDB created by the ISWS served as an excellent starting point when developing a database structure for this project and populating it with data.

Relational Databases

A database is constructed with the data structures such as data objects, the governing rules, and associations related to the data objects based on a concept, a data model. A data model is specific to organization of the data instead of type of operations to be executed or hardware and software used. In this way, a data model correlates the concepts that make up real-world events and processes with the physical representation of those concepts in a database. In addition to being relatively easy to create and access, a relational database has the important advantage of being easy to extend. After the original database creation, a new data category can be added without requiring that all existing applications be modified.

A relational database is a set of tables containing data in predefined categories. Each table contains one or more data categories in columns. Each row contains a unique instance of data for the categories defined by the columns. The tables are then related back to each other by the database engine when requested. A database user can obtain a *view* of the database that fits the user's needs. While creating a relational database, one can define a *domain* of possible values in a data column and further *constraints* that may apply to that data value. The standard user and application program interface to a relational database is the *structured query language* (SQL). SQL statements are used both for interactive queries to retrieve data and for displaying data in reports.

Tables include a unique identifier for each instance. This unique identifier can be used in other tables to refer to the particular instance without repeating all the information about that instance again, thus providing necessary links among related tables. The process of removing redundant data from a relational database by separating information into smaller tables is called normalization. A normalized database, a database with relations that follow a series of rigorous standards, generally improves performance, lowers storage requirements, and makes it easier to change the application or to add new features.

STAR Environmental Database (STARED)

A comprehensive database, *STAR Environmental Database (STARED)* is developed to store various environmental data, including water quality, sediment chemistry, biological indices, stream hydrology, and habitat. The structure is based on a structure of the FoxDB, the relational database developed by the Illinois State Water Survey (McConkey *et al.,* 2004). FoxDB was developed to compile water quality data for the Fox River watershed from a variety of sources. It

contains all available water and sediment quality data collected in the Fox River and its tributaries since 1970s, making it very convenient starting point for development of STARED. The FoxDB structure was further modified and expanded to include raw biological (taxonomic) data and habitat information.

Figure 2 demonstrates the structure of STARED. Colored blocks group tables with a common theme such as sampling stations, samples taken at these sites, monitoring projects, parameters analyzed, results of analyses, and taxonomic information (counterclockwise direction from bottom right corner). Tables are related through arrows based on unique identifiers. Each table within a block then provides attributes describing the theme or providing lookup information. For example, a table *TBLSample* describing a sample collected by a crew at a sampling station includes a sample number uniquely identifying the sample, sampling date and time, sampling depth, medium, sampling stations, monitoring project etc. Another table explains codes used to describe the sampled medium, e.g. “W” as water, “S” as sediment, or “M” as macroinvertebrate taxa, or monitoring project. Codes referring to projects are fully described in a separate table, *TBLProjects_Programs*.

Database maintenance and data import are done with the help of *IDLocations* code that refers to the original file acquired from the particular data source. The table *TBLIDLocations* is not included into any of the above blocks, and is shown separately in the data model.

A station is defined in table *TBLStation_Information* with a unique identifier, “Station ID”, and several descriptive fields. Latitude, longitude, and standard identifiers such as National Hydrography Dataset (NHD) Code and Reach File Version 3 (RF3) Code provide means to display stations in GIS environment and to relate them to national datasets. “Station_Type” identifies whether the station is located on a stream, a lake, in a wetland etc. Codes are explained in a lookup table, *TBLStation_Type*. Additional columns and look up tables include description of the site, waterbody name, EPA or USGS station codes (if relevant), latitude and longitude accuracy level, and contributing area (when available from the original source). Watershed and reach level information derived using GIS can be found in tables related through “Station ID”. Separate table stores flow measurements, currently for selected USGS stations only.

In the Sample Related block, the table *TBLSample* describes a sample with the information of the sampling station, sampling date and time, sampling depth and a monitoring project under which it was collected. *TBLSample* is connected to three look up tables. *TBLMedium* indicates what was sampled (water, sediment, biota, habitat characteristics). *TBLSample_Type* describes sampling methods (transect composite, grab sample, continuous datasonde, fish taxa, etc.). *TBLComposite_statistic_code* indicates whether the measured value is an individual value or an average value (based on STORET database).

The Project Related block is centered on the table *TBLProjects_Programs*, linked to three other tables. *TBLProjects_Programs* contains the records of monitoring project names with descriptions of study areas, project objectives and dates, codes for the monitoring organization, and contact information. *TBLOrganization* consists of full and abbreviated names and category of the organization, including its postal and web addresses. *TBLZip* simplifies recording of the organization postal address.

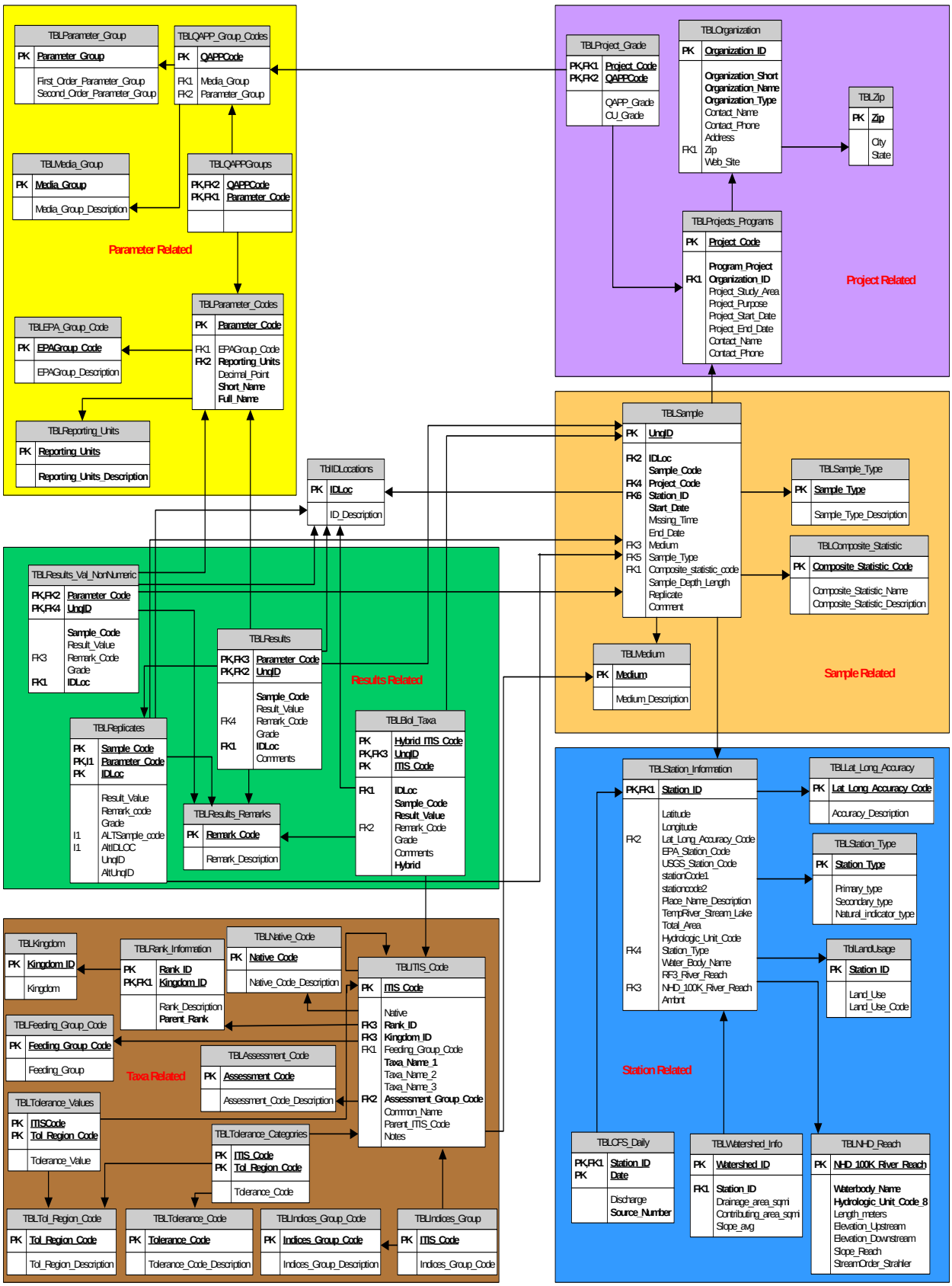

Figure 9 STARED Structure.

Parameter codes are adopted from Legacy STORET. Although the EPA is retreating from using these 6-digit codes, most available data are still referenced this way. New 7-digit codes were created specifically in this project for habitat parameters or biological indices not included in the original STORET codes but needed for the analyses. The table *TBLParameter_Codes* in the block *Parameter Related* closely follows the parameter table from Legacy STORET with full and abbreviated description of parameters, reporting unit, and accuracy. Look up information is provided in the following tables: *TBLReporting_Units*, *TBLGroup_Code*, *TBLMedia_Group*, *TBLParameter_Group*, *TBLGroup_Codes*, and *TBLQAPP_Groups*.

For grouping parameters, two schemes, Legacy STORET scheme and QAPP scheme are used in this database. The QAPP scheme was developed by the ISWS (McConkey *et al.* 2004) together with the QAPP grading system to allow the evaluation of data quality. The QAPP scheme groups parameters on two levels, by sampled medium, and by constituent analyzed. The three-digit coding scheme of QAPP enables to identify the medium, the main parameter group and the constituent subgroup. The main parameter group includes basic inorganic, nutrients, metals and organics; and the constituent subgroup comprises of number of groups such as nitrogen in the nutrients group or pesticides in the organics group.

The *Taxa Related* block provides taxonomic information on aquatic biota, presently fish and macroinvertebrates although the structure enables incorporating other taxonomic groups such as algae or macrophytes. The USDA houses Integrated Taxonomic Information System (IT IS) database developed to provide accurate, scientifically credible, and current taxonomic data and to serve as a standard to enable the comparison of biodiversity datasets (USDA, 2004). The IT IS taxonomic classification and codes were adopted into the STARED. The IT IS code is similar to the STORET parameter codes -- it basically describes what can be found or analyzed in a sample. The taxonomic part of the table *TBLITIS_Code* mimics the IT IS structure defining taxonomic hierarchy with Latin and common names, parent taxa, and taxonomic rank. Other information relating specifically to this project includes the assessment group (fish or macroinvertebrates), and a code specifying whether the species is native. The table *TBLIndices_Group* assigns species or taxa to the most common groups used in deriving index of biotic integrity, such as Amphipods or Chironomids for macroinvertebrate indexes and darters or simple lithophilic spawners for fish indexes. Indices also include feeding preferences of the taxa, e.g., collectors, gatherers, herbivores, or insectivores.

Results described in the *Results Block* define actual values of parameters analyzed in a sample. Numerical and non-numerical results are stored in separate tables, *TBLResults* and *TBLResults_Vol_NonNumeric*, respectively. The third table, *TBLReplicates* is used to store all replicate results. Biological ‘catch’ data are stored in the table *TBLBio_Taxa*. The structure of these tables is very similar. For each sample identified by a unique ID, the result is the concentration value for a parameter specified by the Parameter Code, or number of individuals for a species defined by the IT IS Code, respectively. A remark code may accompany a result with additional information about the quality issues such as “below the detection limits” or “calculated value”. Unreliable or questionable data may be indicated with an optional grade and comment.

The complete data dictionary with description of the tables and fields is given in Appendix.

Database Management and Implementation

The master database is stored in Microsoft SQL Server 2000 format on a server connected to a network. This setup allows multiple users to access the data either directly using SQL Server or the Open Database Connectivity (ODBC) interface. The ODBC interface allows data accessing among various software applications regardless of vendor. For example, the user can link Microsoft Access to tables stored in Microsoft SQL Server and access the data in real time. Tables and queries created in Access or SQL Server can also be easily imported to Excel, ArcGIS or other software for display and analysis.

A two-level access database architecture is recommended. All users can access the database through a client connection and query the database to extract desired information. Individual users do not need their personal copy of the database as they are connected via the network to the master copy. Considering the data security, only the database manager has full access to the database, can add and delete data, and modify the database structure. All other users can forward any relevant and preformatted data to the database manager for import. Their privileges are specified as read-only.

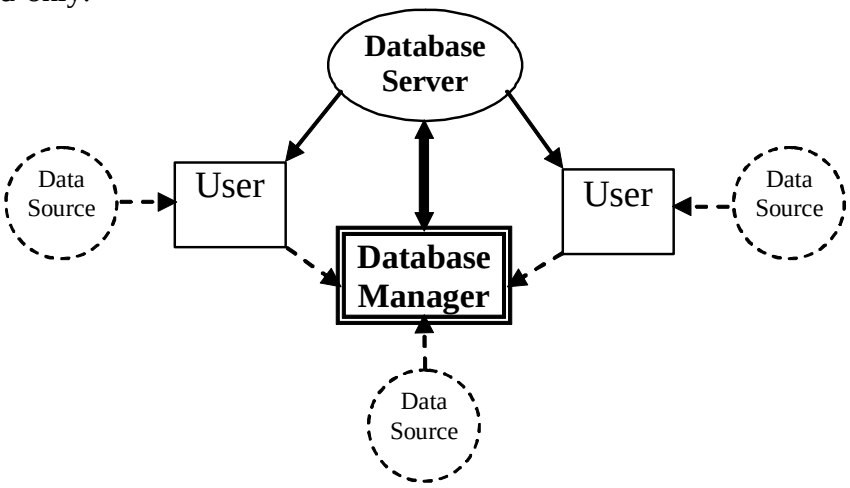

Figure 10 Database architecture, showing multiple users connected to a single database server.

The master copy of the database, *STARED* is currently stored on SQL Server, *SPRUCE18* at Northeastern University, Boston. The data manager operates from Center for Urban and Environmental Studies at Northeastern University to update the database structure, to coordinate data import, to provide necessary quality control, and to ensure database integrity. The personnel in the Information System/System Administration Department of the Northeastern University are responsible for the maintenance of the server. User from both within and outside the NEU network can access the database.

Data acquired and preformatted by the project team are forwarded to the data manager for final quality check and import. Any documentation aiding in interpreting the data beyond the information stored in STARED is saved on *SPRUCE18* in a separate folder.

Data Sources and Availability

The developed database is targeted to contain data from a variety of sources. Data acquired by the ISWS in the FoxDB (McConkey *et al.* 2004) represent an integral part of STARED. Additional data were acquired from major federal and state agencies collecting data in the study area.

These agencies include the U.S. Environmental Protection Agency (USEPA), Ohio Environmental Protection Agency (OEPA), Illinois Department of Natural Resources (Illinois DNR), Illinois Environmental Protection Agency (IEPA), Maryland Department of Natural Resources (Maryland DNR), U.S. Geological Survey (USGS), Minnesota Pollution Control Agency (MPCA) and the Massachusetts Department of Environmental Protection (MADEP).

Federal Sources

The USEPA STOrage and RETrieval (STORET) System, the USGS National Water-Quality Assessment (NAWQA) Program, and the USGS National Water Information System (NWIS) are major federal databases of environmental data available on the internet.

STORET Data

The STORET is a computerized database containing data on physical characteristics and chemical constituents of water, fish tissue, and sediment. It also includes information on municipal waste sources and disposal systems, data on pollution-caused fish kills, daily stream flow, and geographic and other descriptive data for collection sites. The US EPA maintains two data management systems containing water quality information for the Nation's waters: the STORET Legacy Data Center (LDC) and the modernized STORET. The LDC contains historical water quality data dating back from the early part of the 20th century to the end of 1998. The STORET system contains data beginning in 1999, along with older data that have been properly documented and migrated from the LDC (USEPA, 2005).

Over the years, STORET has evolved into a fully decentralized database and all responsibility for its operation and maintenance resides with the end users, who each maintain a customized copy of their data. The STORET software, designed to make local manipulation of these data manageable for the user, is installed on top of each user's customized database. In the STORET architecture, the EPA provides the technical design of the database and controls which fields exist as well as the field properties. The EPA provides the data to initialize certain tables. The users are responsible for providing the Oracle database environment within which STORET resides. The users upload their data regularly to the centralized warehouse open to public access.

The STORET uses the data warehouse concept that allows the general public to access data via the Internet. A data warehouse is an aggregation of related data, maintained centrally, and easily accessed by a broad community of users with minimal skills. Under this concept the EPA maintains a centrally located copy of all the data submitted to the EPA from the STORET servers around the country by partner clients (USEPA, 2003).

USGS/NAWQA Data

Water quality data from the USGS are available through the NWIS. The NWIS is a distributed database in which data can be processed over a network of workstations and fileservers at the USGS offices throughout the United States. The system has four components, including the Water Quality System, which is the most relevant for the current project. The Water Quality System contains the results of more than 3.5 million analyses of water samples that describe the chemical, physical, biological, and radiochemical characteristics of both surface water and groundwater. The Web site provides current and historical data (USGS, 2005a).

Data from NAWQA are stored in a separate database (USGS, 2005b). The USGS began its NAWQA program in 1991, systematically collecting chemical, biological, and physical water quality data from study units across the nation. The first cycle, which was completed in 2001, involved 59 study units that covered about 65 percent of water used for drinking and irrigation. For the second cycle, 13 study units were eliminated and 8 were combined into 4 study units, leaving 42 study units that still cover more than 60 percent of the national water uses.

The Fox River Watershed Water Quality Database (FoxDB)

The FoxDB is a water quality database developed by the ISWS during the Fox River Watershed Investigation, Phase I. All available data on water quality (water and sediment chemistry data) and other related parameters that define the nature of the stream and river environment were compiled into one database. Stream flow data are included as an integral part to interpret reported concentrations of chemical water constituents. This database has been designed so that it can be expanded in the future to include other types of data and data from other watersheds (McConkey *et al.,* 2004).

The FoxDB contains the data that have been collected over the years by various agencies and groups as a result of a variety of monitoring activities. These include long-term as well as short-term projects. Regular monitoring programs of the U.S. Environmental Protection Agency (USEPA), Illinois Environmental Protection Agency (IEPA), and the U.S. Geological Survey (USGS) represent a major portion of data available for the watershed. These data were acquired from the USEPA Legacy Data Center (formerly STORET), the USGS National Water Information System (NWIS), and the USGS National Ambient Water Quality Assessment (NAWQA) databases (all available online).

In addition to the regular monitoring programs described above, data from several short-term studies were used in populating the FoxDB. The Upper Illinois Basin Urban Gradient Study has been conducted by the USGS under NAWQA study program in years 1999-2001. Though this study is not a part of the FoxDB, it also provides data describing habitat, fish, periphyton, and microinvertebrate populations for the Upper and Lower Fox River. All biological data include taxonomic identifications to the lowest practical taxonomic level (genus or species) (Adolphson *et al.,* 2002). A two-year study conducted by Max McGraw Wildlife Foundation (MMGWF) investigated the environmental effects of dams on fisheries in a 100-mile stretch of the Fox River between the Fox Chain of Lakes and Dayton, Illinois (Santucci and Gephard, 2003). Sediment, water, fish, and macroinvertebrate samples were taken at 40 sites located in a free-flowing river directly below dams, impounded river directly above dams, and free-flowing or impounded mid-

segment areas between dams during summer low flow periods. Details on these and other datasets included in the FoxDB can be found in McConkey *et al.* (2004).

State Sources

Maryland Biological Stream Survey (MBSS) Data
Maryland Biological Stream Survey data includes 955 first, second and third-order stream segments, encompassing all 17 major drainage basins in the state of Maryland over the three-year sampling period (1995-1997). Statewide and basinwide results and an assessment of the condition of the streams were reported in the MBSS three-year report (Roth *et al.*, 1999). Water chemistry and benthic macroinvertebrates were analyzed in spring (March-April) while fish, physical habitat, and in-situ water chemistry were analyzed in summer (June-September). All sampling sites are classified into three geographic regions: West, Central, and East. Biological measurements include abundance and health of fish, composition of benthic macroinvertebrate communities, and presence of amphibians and reptiles, aquatic plants, and mussels. Chemical measurements include pH, sulfate, nitrate-nitrogen, conductivity, and dissolved oxygen. Physical habitat measurements took into account parameters such as flow, stream gradient, maximum depth, embeddedness, instream habitat, epifaunal substrate, pool and riffle quality, bank stability, channel flow status, shading, and riparian buffer type (Mercurio *et al.*, 1999).

Ohio EPA Data
The Ohio dataset was assembled from the chemical, habitat and biological data collected by the Ohio EPA since 1967. However, the data available for the period before 1990 are few in numbers. The chemical data are available for water, sediment, and fish tissue. The original data set is available in the FoxPro format. The fish tissue database currently holds 5,058 samples collected from 1967 through 1996. The fish tissue was analyzed for pesticides and PCBs (3,978 samples), metals (2,865 samples), VOCs (57 samples), BNAs (166 samples) and herbicides (44 samples).

Minnesota Pollution Control Agency Data
The Minnesota Pollution Control Agency (MPCA) biological data are acquired, and reformatted to fit STARED structure (MPCA, 2005; Genet and Chirhart, 2004). This data set consists of data spread for the whole state, and of twenty year period data.

Illinois EPA Data
The Illinois EPA conducts a wide variety of water quality monitoring programs. Stations are sampled for biological, chemical and/or in-stream habitat data, as well as streamflow. Fixed network of stations is sampled on a 6-week sampling frequency with the samples analyzed for a minimum of 55 universal parameters, including field pH, temperature, specific conductance, dissolved oxygen (DO), suspended solids, nutrients, fecal coliform bacteria, and total and dissolved heavy metals (IEPA, 2005). The monitoring program also includes intensive stream surveys (incl. biological and habitat data) with all watersheds being sampled once in a 5-year rotation.

Water chemistry data for the Fox River watershed were already a part of the FoxDB. Biological and habitat data were acquired from the Illinois EPA and imported into STARED.

Intensive river basin surveys are conducted on a five-year rotational basis in cooperation with the Illinois Department of Natural Resources (IDNR). These intensive surveys are a major source of information for annual 305(b) assessments. Water chemistry and biological data (fish and macroinvertebrates) and qualitative and quantitative in-stream habitat information, including stream discharge, are collected to characterize stream segments within the basin, identify water quality conditions, and evaluate aquatic life use impairment. Fish tissue contaminant and sediment chemistry sampling also are conducted to screen for the accumulation of toxic substances (IEPA, 2005).

QA/QC Procedure

Attribute Assignment
Each dataset comes with its own set of attributes assigned by a data owner. These attributes need to be translated into the attributes and codes used in STARED. Translation tables for the federal sources of data were developed by McConkey *et al.* (2004).

Both the LDC and NWIS, unlike STORET, use the same parameter codes. These five-digit codes (Table 5) are used to identify a measured constituent that also embeds information on the units, medium, and procedures. Data from other sources were reviewed and appropriate five-digit parameter codes assigned.

Table 5 Example of five-digit parameter codes used in the NWIS and LDC (USGS 2004b).

Code	Parameter Definition
00598	Dissolved nitrogen gas, water, unfiltered, percent of saturation
00600	Total nitrogen, water, unfiltered, milligrams per liter
00601	Total nitrogen, suspended sediment, total, milligrams per liter
00602	Total nitrogen, water, filtered, milligrams per liter
00603	Total nitrogen, bed sediment, total, dry weight, milligrams per kilogram
00604	Ammonia, suspended sediment, total, milligrams per liter as nitrogen
00605	Organic nitrogen, water, unfiltered, milligrams per liter

Removing Duplicates
The data for each watershed are obtained from several different sources. Thus, data have to be checked for duplicates (*i.e.*, samples taken at the same place and time analyzed for the same parameters, but included in two or more databases). Most apparent duplicate results originate from overlaps between databases. It has been mentioned that USGS maintains two water quality databases, NWIS and NAWQA. According to McConkey *et al.* (2004), most of the data in the NWIS database represent rounded values of data in the NAWQA database. However, distinct data exist in each database. Similarly, some duplicates exist between the LDC and the USGS data. Duplicate entries were moved to a separate table *TblReplicates* structured after the results table *TblResults*.

True replicate samples appear from Quality Assurance and Quality Control (QA/QC) procedures. A sampling crew or a laboratory may be required to take multiple samples or analyze a sample multiple times at a frequency specified in a Quality Assurance Project Plan (QAPP). These samples are marked as replicates and moved to table *TblReplicates*.

Station Location
Location of monitoring sites defined by latitude and longitude coordinates was displayed in GIS environment and compared against a detailed site description. Geographical coordinates were adjusted to match the description as needed and latitude – longitude accuracy code changed to reflect the adjustment.

Stations in FoxDB were already inspected by the ISWS. Monitoring sites from new data were first checked against stations already stored in the database. Location of any new station was verified and corrected when necessary. Additional information such as topographic maps, hydrography, or roads was used to assist in determining correct location.

Spatial Data – GIS

Spatial data is formatted to be displayed and analyzed in Geographic Information System software. Examples of spatial data formats include: digital elevation models (DEM) stored as raster data, ArcGIS layer files (or ArcView shapefiles) representing monitoring point locations, land use, ecoregions, and soil type, and hydrography files describing the shape and spatial properties of streams. Many federal and state agencies operate and maintain databases of spatial information, including downloadable data for use in GIS software.

National Hydrography Dataset (NHD) is developed on Digital Line Graph (DLG) of USGS integrating the Reach File Version 3 (RF3) of USEPA to provide information on waterbodies such as rivers, ponds, streams and lakes. The NHD supersedes RF3 and DLG datasets. The NHD can be downloaded in three different resolutions, high, medium and local. Medium resolution (1:100000) is available for the conterminous States area. High (1:24000) or local (varies) resolution hydrography is being developed and its availability varies among the States and watersheds. The full description of the NHD data as well as a download tool can be found at http://nhd.usgs.gov/

The NRCS provides 1:250,000 scale digital soil information from the State Soil Geographic Database (STATSGO). This digital geographic data is available in several formats including: digital line graph files and ARC/INFO coverages. The STATSGO includes information about the location of soil types and are linked to the Soil Interpretations Record (SIR) attribute database. The SIR includes information about soils' respective properties, including 25 physical and chemical properties. Higher resolution data (SSURGO) may be available on a county level.

The most current information on land use can be found on a state level. Statewise GIS coverages on Illinois state land use information can be obtained from the Illinois Department of Agriculture. The Illinois Interagency Landscape Classification Project (IILCP) produced a coverage detailing land use in 1999–2000. The primary source for this digital information was LANDSAT satellite imagery from three different seasons and is classified by 23 different land use categories. Wisconsin state land use information can be obtained from the Wisconsin Department of Natural Resources (Wisconsin DNR) and the Wisconsin Initiative for Statewide Cooperation on Landscape Analysis and Data (WISCLAND). The source for these data was LANDSAT Thematic Mapper satellite imagery. The land use data is organized by 38 hierarchical classifications. The data are available for download in ArcInfo Grid and TIF formats (WDNR, 1999). Land use data for the states of Maryland, Ohio and Minnesota can also be obtained from the DNR of each state. (For Maryland, at http://dnrweb.dnr.state.md.us/gis/data/, for Ohio, at http://www.dnr.state.oh.us/water/gismain/ and for Minnesota, at http://deli.dnr.state.mn.us/)

Ecoregion GIS coverages for the interested regions in United States of America were downloaded from EPA at http://www.epa.gov/wed/pages/ecoregions/level_iii.htm. National Inventory of Dams (NID) data compiled by US Army Corps of Engineers are downloaded from http://crunch.tec.army.mil/nid/webpages/nid.cfm.

Data Query Example

The database is very flexible, allowing a retrieval of data in various forms, tables, or summaries. This section shows several examples of queries designed in MS Access, corresponding SQL statements as well as resulting tables. Examples include queries on both water quality and biological data, extracting the raw data as well as creating summaries.

Data Retrieval

Query 1: Find result values for a particular constituent and station
The query shown in Figure 11 retrieves all copper measurements at a selected station. To limit the number of results further, a constraint has been also specified for a monitoring year. The tables needed for executing this query are: TBLSample, TBLResults, and TBLParameter_Codes. Additional useful information can be gained from including tables TBLMedium, TBLProjects_Programs, and TBLOrganization.

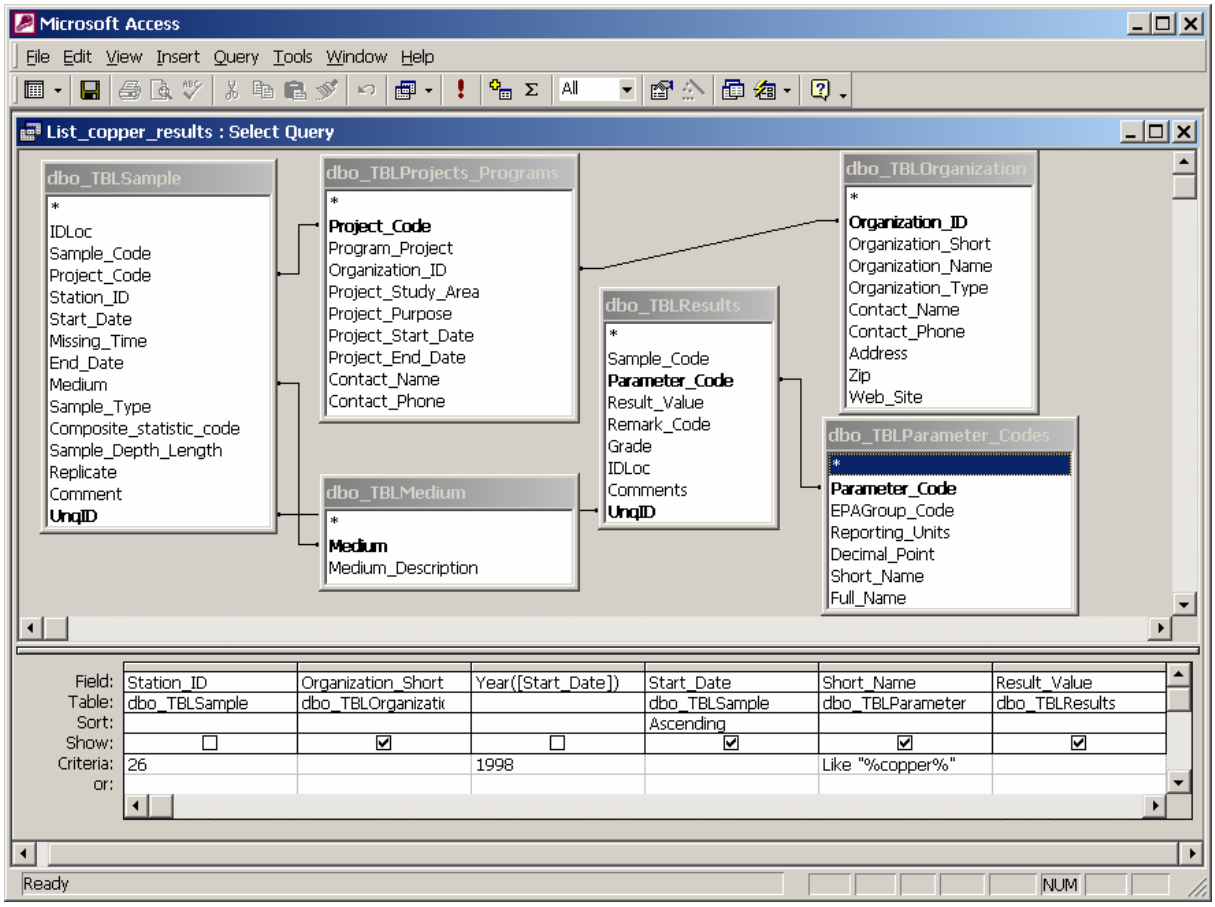

Figure 11 List stations on the Fox River sampled for copper concentration after 1995.

The corresponding SQL code for running this query is:

```
SELECT dbo_TBLOrganization.Organization_Short, dbo_TBLSample.Start_Date,
dbo_TBLParameter_Codes.Short_Name, dbo_TBLResults.Result_Value,
dbo_TBLResults.Remark_Code, dbo_TBLMedium.Medium_Description
```

```
FROM ((((dbo_TBLSample INNER JOIN dbo_TBLResults ON dbo_TBLSample.UnqID =
dbo_TBLResults.UnqID) INNER JOIN dbo_TBLParameter_Codes ON
dbo_TBLResults.Parameter_Code = dbo_TBLParameter_Codes.Parameter_Code) INNER JOIN
dbo_TBLMedium ON dbo_TBLSample.Medium = dbo_TBLMedium.Medium) INNER JOIN
dbo_TBLProjects_Programs ON dbo_TBLSample.Project_Code =
dbo_TBLProjects_Programs.Project_Code) INNER JOIN dbo_TBLOrganization ON
dbo_TBLProjects_Programs.Organization_ID = dbo_TBLOrganization.Organization_ID
WHERE (((dbo_TBLSample.Station_ID)=26) AND ((Year([Start_Date]))=1998) AND
((dbo_TBLParameter_Codes.Short_Name) Like "%copper%"))
ORDER BY dbo_TBLSample.Start_Date;
```

Table 6 shows the query results sorted by date. Two organizations sampled at the selected station during 1998: Illinois Environmental Protection Agency (IEPA) and Fox River Water Reclamation District (FRWRD). The IEPA analyzed the sample for both total and dissolved concentration while the FRWRD determined only total copper concentrations in the sample. The detection limit of the method used by the IEPA is not low enough to detect the ambient concentration. The FRWRD uses a more precise method able to quantify the copper concentration. There was no sediment sample tested for copper at this station in 1998.

Table 6 Copper concentration measured at Station 26 during 1998.

Organization	Start_Date	Short_Name	Result_Value	Remark_Code	Medium_Description
IEPA	02/04/1998 1:22:00	COPPER CU,DISS	10	K	Water
IEPA	02/04/1998 1:22:00	COPPER CU,TOT	10	K	Water
IEPA	03/17/1998 12:45:00	COPPER CU,TOT	10	K	Water
IEPA	03/17/1998 12:45:00	COPPER CU,DISS	10	K	Water
IEPA	04/09/1998 12:46:00	COPPER CU,DISS	10	K	Water
IEPA	04/09/1998 12:46:00	COPPER CU,TOT	10	K	Water
IEPA	06/01/1998 1:22:00	COPPER CU,TOT	10	K	Water
IEPA	06/01/1998 1:22:00	COPPER CU,DISS	10	K	Water
IEPA	07/06/1998 2:05:00	COPPER CU,DISS	10	K	Water
IEPA	07/06/1998 2:05:00	COPPER CU,TOT	10	K	Water
IEPA	08/10/1998 2:00:00	COPPER CU,TOT	10		Water
IEPA	08/10/1998 2:00:00	COPPER CU,DISS	10	K	Water
FRWRD	08/19/1998	COPPER CU,TOT	9		Water
FRWRD	09/02/1998	COPPER CU,TOT	5.8		Water
IEPA	09/15/1998 1:00:00	COPPER CU,DISS	10	K	Water
IEPA	09/15/1998 1:00:00	COPPER CU,TOT	10	K	Water
FRWRD	09/16/1998	COPPER CU,TOT	7.3		Water
FRWRD	09/23/1998	COPPER CU,TOT	6.9		Water
FRWRD	09/30/1998	COPPER CU,TOT	7.3		Water
FRWRD	10/28/1998	COPPER CU,TOT	8		Water
FRWRD	11/04/1998	COPPER CU,TOT	8		Water
IEPA	11/12/1998 2:06:00	COPPER CU,DISS	10	K	Water
IEPA	11/12/1998 2:06:00	COPPER CU,TOT	10	K	Water
FRWRD	11/18/1998	COPPER CU,TOT	6		Water
FRWRD	12/02/1998	COPPER CU,TOT	5.4		Water
IEPA	12/07/1998 12:20:00	COPPER CU,TOT	10	K	Water
IEPA	12/07/1998 12:20:00	COPPER CU,DISS	10	K	Water

Note: K → Actual value not known, but known to be less than value shown.

Summaries

Query 2: List stations on selected stream sampled for a particular constituent
The query shown in Figure 12 generates a list of stations with samples analyzed for copper. To limit the number of results further, a constraint has been specified for a water body name and year. The results include the information of number of samples of any listed stations. The tables needed for executing this query are: TBLStation_Information, TBLSample, TBLResults, TBLParameter_Codes, and TBLMedium

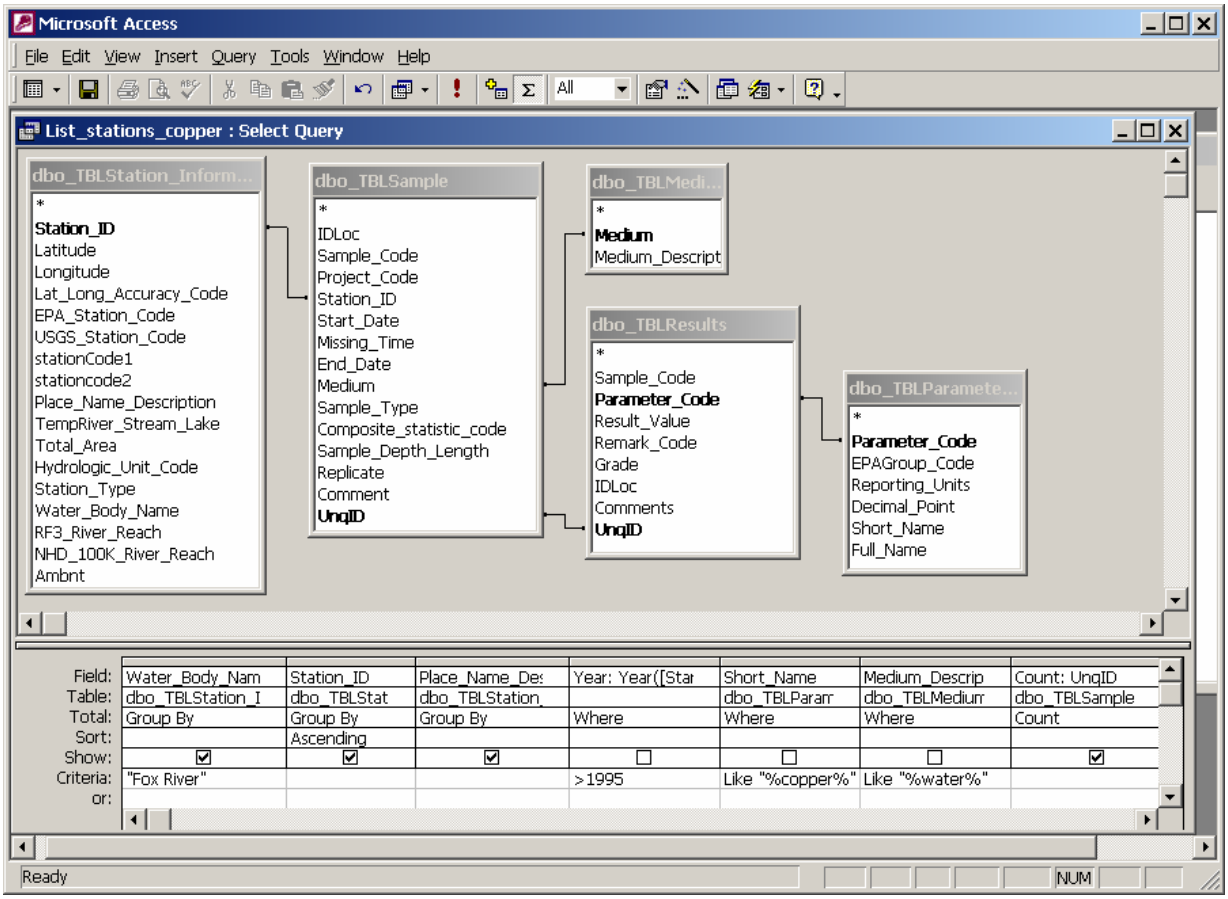

Figure 12 List stations on the Fox River sampled for copper concentration after 1995.

The corresponding SQL code for running this query is:

```
SELECT dbo_TBLStation_Information.Water_Body_Name,
dbo_TBLStation_Information.Station_ID,
dbo_TBLStation_Information.Place_Name_Description, Count(dbo_TBLSample.UnqID) AS
[Count]
FROM (((dbo_TBLSample INNER JOIN dbo_TBLResults ON dbo_TBLSample.UnqID =
dbo_TBLResults.UnqID) INNER JOIN dbo_TBLParameter_Codes ON
dbo_TBLResults.Parameter_Code = dbo_TBLParameter_Codes.Parameter_Code) INNER JOIN
dbo_TBLStation_Information ON dbo_TBLSample.Station_ID =
dbo_TBLStation_Information.Station_ID) INNER JOIN dbo_TBLMedium ON
dbo_TBLSample.Medium = dbo_TBLMedium.Medium
```

```
WHERE (((Year([Start_Date]))>1995) AND ((dbo_TBLParameter_Codes.Short_Name) Like
"%copper%") AND ((dbo_TBLMedium.Medium_Description) Like "%water%"))
GROUP BY dbo_TBLStation_Information.Water_Body_Name,
dbo_TBLStation_Information.Station_ID,
dbo_TBLStation_Information.Place_Name_Description
HAVING (((dbo_TBLStation_Information.Water_Body_Name)="Fox River"))
ORDER BY dbo_TBLStation_Information.Station_ID;
```

Table 7 shows the query results sorted by the station identification number, Station_ID. The number of measurements available for each station is included. The number of measurements is higher than the number of samples taken at the station during the same period as one sample can be analyzed for several forms of copper. For example, if only one sample was taken but copper concentration was determined in both total and dissolved forms then the number of measurements (Count) is two.

Table 7 Stations on the Fox River mainstem with copper concentrations after 1995.

Water_Body_Name	Station_ID	Place_Name_Description	Count
Fox River	23	ROUTE 176 BRUTONS BRIDGE 5 MILES	128
Fox River	24	ROUTE 62 -ALGONQUIN ROAD BRIDGE	126
Fox River	26	STATE ST ELGIN, IL (T41N R8E NW35)	153
Fox River	27	MILL STREET (Old Rt 30) BRIDGE	170
Fox River	30	COUNTY HIGHWAY 18 AT DAYTON, IL	38
Fox River	31	RT 71-6 NE EDGE OTTAWA	82
Fox River	33	RT 34 WASHINGTON STREET BRIDGE OSWEGO	43
Fox River	35	NATIONAL STREET BRIDGE, ELGIN	5
Fox River	37	COUNTY ROAD BRIDGE AT WEDRON	6
Fox River	40	2 MI S of GENEVA FABYAN PK	6
Fox River	162	BETWEEN ELGIN AND SOUTH ELGIN, IL	5
Fox River	197	FOX R RT 173 BR 4 MI E ANTIOCH	126
Fox River	240	FOX R I-90 BR N OF ELGIN	17
Fox River	273	Kimball St Dam N Rt 20/ LAWRENCE ST ELGIN T41N	1

Query 3: List and count fish species in selected stream and year

The query shown in Figure 13 generates a list of fish species found in the Fox River mainstem in a particular year, provides a total number caught for each species regardless of a monitoring organization, and lists the species in descending order based on numbers caught. Fish identified as hybrids are listed and counted separately. This query summarizes the data for all stations on the Fox River sampled during 2000. The tables needed for executing this query are: TBLSample, TBLBio_Taxa, TBLITIS_Code, and TBLStation_Information.

The SQL code corresponding to this query is:

```
SELECT dbo_TBLITIS_Code.Common_Name,
dbo_TBLStation_Information.Water_Body_Name, Year([Start_Date]) AS Expr1,
Count(dbo_TBLBiol_Taxa.ITIS_Code) AS CountOfITIS_Code,
dbo_TBLITIS_Code_1.Common_Name AS Hybrid
FROM (((dbo_TBLBiol_Taxa INNER JOIN dbo_TBLITIS_Code ON
dbo_TBLBiol_Taxa.ITIS_Code = dbo_TBLITIS_Code.ITIS_Code) INNER JOIN
dbo_TBLSample ON dbo_TBLBiol_Taxa.UnqID = dbo_TBLSample.UnqID) INNER JOIN
dbo_TBLStation_Information ON dbo_TBLSample.Station_ID =
```

```
dbo_TBLStation_Information.Station_ID) INNER JOIN dbo_TBLITIS_Code AS
dbo_TBLITIS_Code_1 ON dbo_TBLBiol_Taxa.Hybrid_ITIS_Code =
dbo_TBLITIS_Code_1.ITIS_Code
GROUP BY dbo_TBLITIS_Code.ITIS_Code, dbo_TBLITIS_Code.Common_Name,
dbo_TBLStation_Information.Water_Body_Name, Year([Start_Date]),
dbo_TBLBiol_Taxa.Hybrid_ITIS_Code, dbo_TBLITIS_Code_1.Common_Name
HAVING (((dbo_TBLStation_Information.Water_Body_Name)="Fox River") AND
((Year([Start_Date]))=2000))
ORDER BY Count(dbo_TBLBiol_Taxa.ITIS_Code) DESC;
```

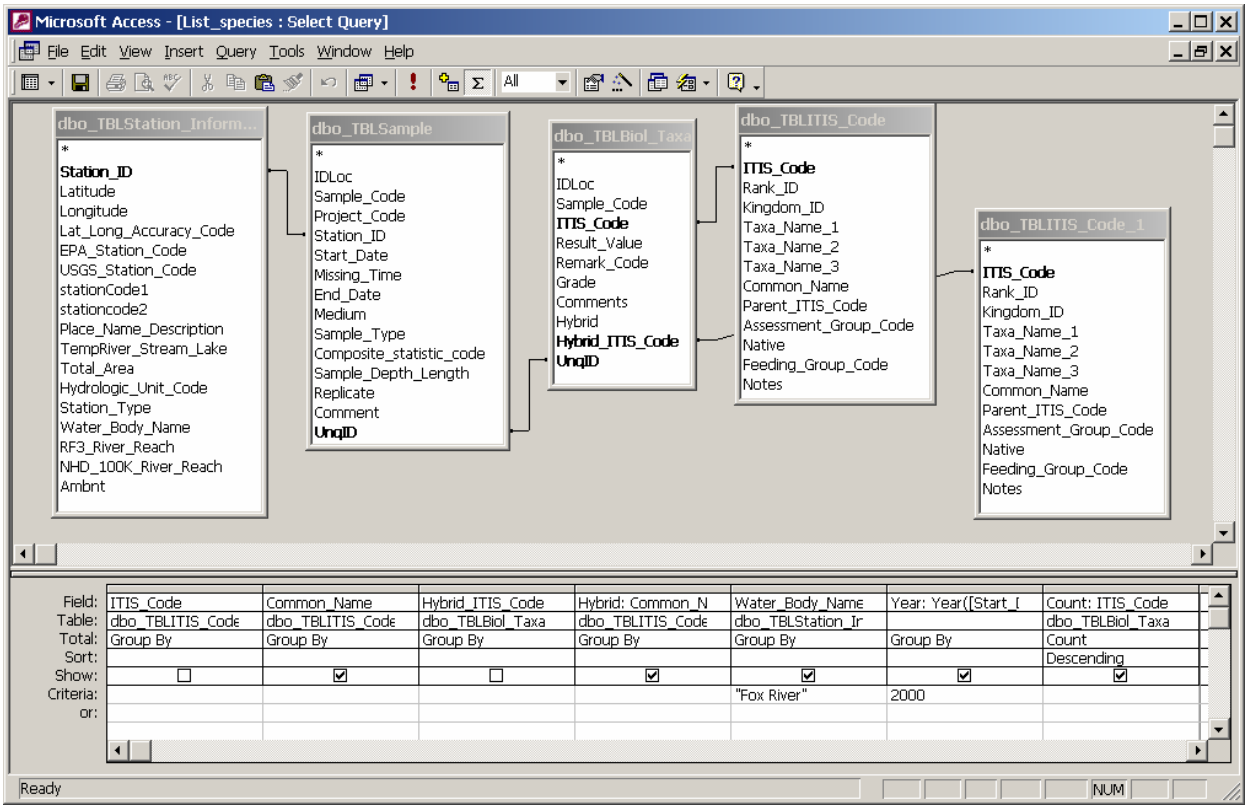

Figure 13 List and count fish species

Table 8 shows the query results sorted by the number of individuals caught during the selected year. Hybrids are identified by two common names – both names of species or taxa are included. Common carp, bluegill, channel catfish, smallmouth bass and largemouth bass are the most common fish species found in the Fox River.

Table 8 Number of fish species found in the Fox River (IL) mainstem in 2000.

Common_Name	Hybrid Species	Water_Body_Name	Year	Count
common carp	NOT HYBRID	Fox River	2000	40
bluegill	NOT HYBRID	Fox River	2000	36
channel catfish	NOT HYBRID	Fox River	2000	35
smallmouth bass	NOT HYBRID	Fox River	2000	30
largemouth bass	NOT HYBRID	Fox River	2000	30
freshwater drum	NOT HYBRID	Fox River	2000	27
flathead catfish	NOT HYBRID	Fox River	2000	27
green sunfish	NOT HYBRID	Fox River	2000	25

spotfin shiner	NOT HYBRID	Fox River	2000	25
quillback	NOT HYBRID	Fox River	2000	21
golden redhorse	NOT HYBRID	Fox River	2000	21
shorthead redhorse	NOT HYBRID	Fox River	2000	15
orangespotted sunfish	NOT HYBRID	Fox River	2000	14
black crappie	NOT HYBRID	Fox River	2000	13
northern hogsucker	NOT HYBRID	Fox River	2000	13
silver redhorse	NOT HYBRID	Fox River	2000	10
bluegill	green sunfish	Fox River	2000	10
walleye	NOT HYBRID	Fox River	2000	10
white bass	NOT HYBRID	Fox River	2000	7
yellow bass	NOT HYBRID	Fox River	2000	7
yellow bullhead	NOT HYBRID	Fox River	2000	7
white sucker	NOT HYBRID	Fox River	2000	7
bluntnose minnow	NOT HYBRID	Fox River	2000	6
logperch	NOT HYBRID	Fox River	2000	6
bullhead minnow	NOT HYBRID	Fox River	2000	5
gizzard shad	NOT HYBRID	Fox River	2000	5
golden shiner	NOT HYBRID	Fox River	2000	4
sand shiner	NOT HYBRID	Fox River	2000	4
river redhorse	NOT HYBRID	Fox River	2000	4
black redhorse	NOT HYBRID	Fox River	2000	4
banded darter	NOT HYBRID	Fox River	2000	4
yellow perch	NOT HYBRID	Fox River	2000	4
slenderhead darter	NOT HYBRID	Fox River	2000	3
spottail shiner	NOT HYBRID	Fox River	2000	3
hornyhead chub	NOT HYBRID	Fox River	2000	2
grass pickerel	NOT HYBRID	Fox River	2000	2
muskellunge	NOT HYBRID	Fox River	2000	2
rock bass	NOT HYBRID	Fox River	2000	2
black bullhead	NOT HYBRID	Fox River	2000	1
smallmouth buffalo	NOT HYBRID	Fox River	2000	1
river carpsucker	NOT HYBRID	Fox River	2000	1
mooneye	NOT HYBRID	Fox River	2000	1
longnose gar	NOT HYBRID	Fox River	2000	1
shortnose gar	NOT HYBRID	Fox River	2000	1
sauger	NOT HYBRID	Fox River	2000	1
blackside darter	NOT HYBRID	Fox River	2000	1
johnny darter	NOT HYBRID	Fox River	2000	1
white crappie	NOT HYBRID	Fox River	2000	1

Calculation of IBI metrics

The structure of STARED is designed to help with calculating metrics describing fish and macroinvertebrate communities in indexes of biotic integrity. The most common metrics are incorporated in the database: table TBLITIS_Code includes information on feeding preferences and whether the species is native or not. The first example calculates number of native and non-native fish taxa caught in the Fox River mainstem in 2000. The second example calculates all available indices.

Query 4: Count native and non-native fish species

The query shown in Figure 14 uses the following tables: TBLSample, TBLBiol_Taxa, TBLITIS_Code, and TBLNative_Code

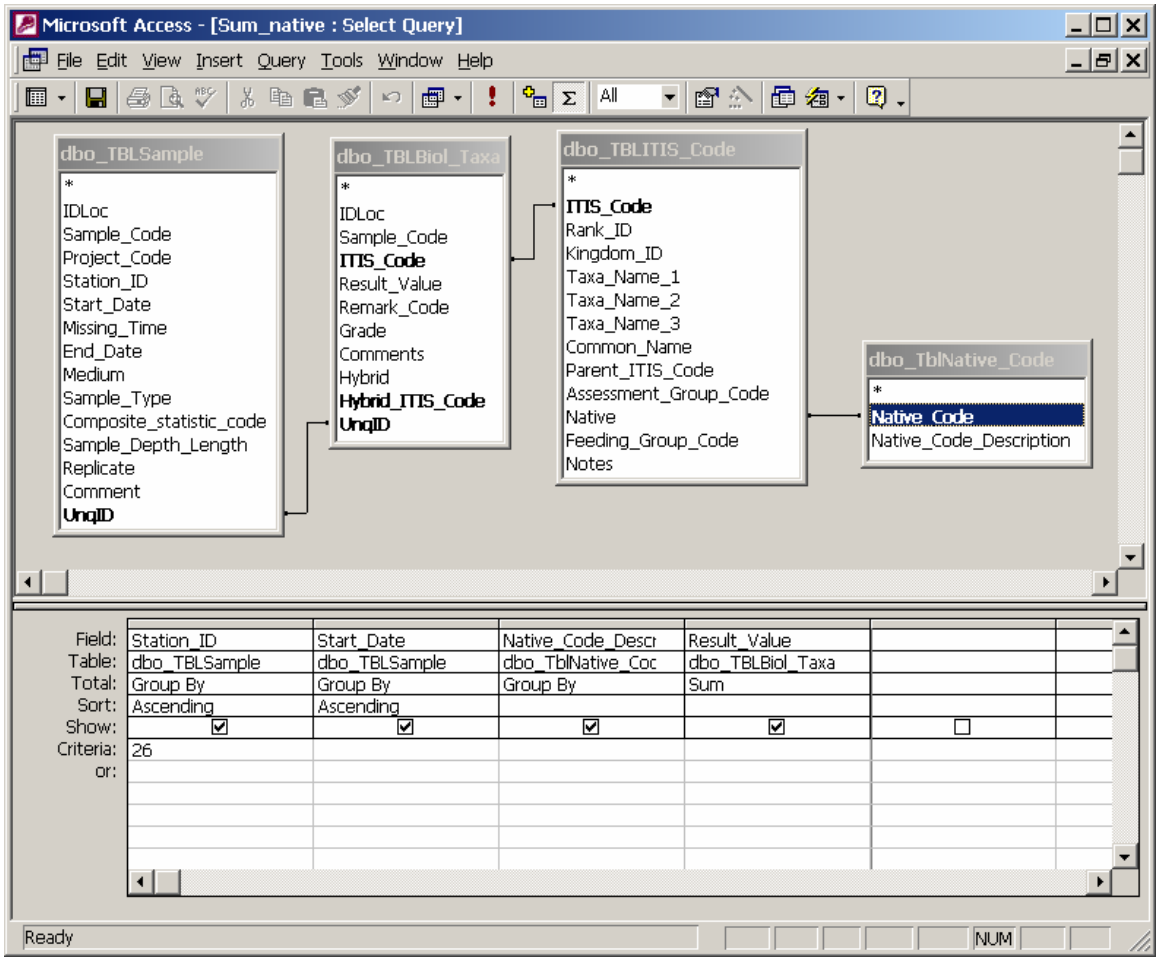

Figure 14 Query 3 – Count native and non-native fish species

The corresponding SQL code is:

```
SELECT dbo_TBLSample.Station_ID, dbo_TBLSample.Start_Date,
dbo_TblNative_Code.Native_Code_Description, Sum(dbo_TBLBiol_Taxa.Result_Value) AS
SumOfResult_Value
FROM dbo_TblNative_Code INNER JOIN ((dbo_TBLBiol_Taxa INNER JOIN
dbo_TBLITIS_Code ON dbo_TBLBiol_Taxa.ITIS_Code = dbo_TBLITIS_Code.ITIS_Code)
INNER JOIN dbo_TBLSample ON dbo_TBLBiol_Taxa.UnqID = dbo_TBLSample.UnqID) ON
dbo_TblNative_Code.Native_Code = dbo_TBLITIS_Code.Native
GROUP BY dbo_TBLSample.Station_ID, dbo_TBLSample.Start_Date,
dbo_TblNative_Code.Native_Code_Description
ORDER BY dbo_TBLSample.Station_ID, dbo_TBLSample.Start_Date;
```

Table 9 shows the query results sorted by the station identification number. Only first 10 results are shown here as the full query contains 40 stations, i.e. 80 results. A sum of native and non-native species will give a total number of species caught at the station. This query includes all

identified fish. Hybrids are counted under the same category as the main species. Adding a constraint on Hybrid_IT IS_Code=0 would limit the query and exclude all hybrids.

Table 9 Number of native and non-native fish species, partial results.

Station_ID	Start_Date	Native_Code_Description	SumOfResult_Value
26	07/27/2000	Native species/taxa	47
26	07/27/2000	Non-native species/taxa	39
33	08/28/2000	Native species/taxa	55
33	08/28/2000	Non-native species/taxa	35
70	09/01/2000	Native species/taxa	90
70	09/01/2000	Non-native species/taxa	14
90	08/18/2000	Native species/taxa	16
90	08/18/2000	Non-native species/taxa	48
259	07/17/2000	Native species/taxa	29
259	07/17/2000	Non-native species/taxa	19

Query 5: Calculate biological indices

This query calculates all available community indices for a sample taken at a specified station and date (). The following tables are required: TBLSample, TBLBiol_Taxa, TBLITIS_Code, TBLIndices_Group, and TBLIndices_Group_Description.

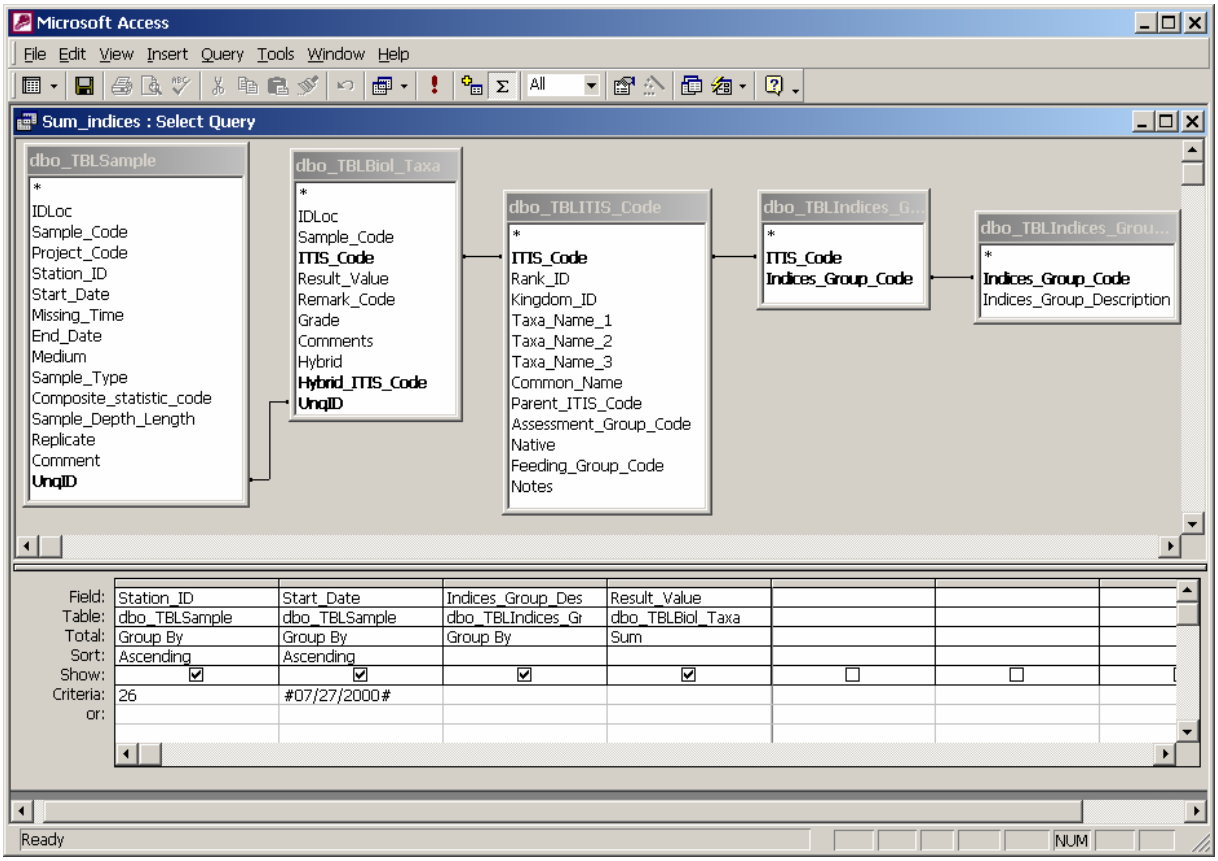

Figure 15 Calculate biological indices for Station 26 on 7/27/2000.

The corresponding SQL statement is:

```
SELECT dbo_TBLSample.Station_ID, dbo_TBLSample.Start_Date,
dbo_TBLIndices_Group_Code.Indices_Group_Description,
Sum(dbo_TBLBiol_Taxa.Result_Value) AS SumOfResult_Value
FROM (((dbo_TBLBiol_Taxa INNER JOIN dbo_TBLITIS_Code ON
dbo_TBLBiol_Taxa.ITIS_Code = dbo_TBLITIS_Code.ITIS_Code) INNER JOIN
dbo_TBLIndices_Group ON dbo_TBLITIS_Code.ITIS_Code =
dbo_TBLIndices_Group.ITIS_Code) INNER JOIN dbo_TBLSample ON
dbo_TBLBiol_Taxa.UnqID = dbo_TBLSample.UnqID) INNER JOIN
dbo_TBLIndices_Group_Code ON dbo_TBLIndices_Group.Indices_Group_Code =
dbo_TBLIndices_Group_Code.Indices_Group_Code
GROUP BY dbo_TBLSample.Station_ID, dbo_TBLSample.Start_Date,
dbo_TBLIndices_Group_Code.Indices_Group_Description
HAVING (((dbo_TBLSample.Station_ID)=26) AND
((dbo_TBLSample.Start_Date)=#7/27/2000#))
ORDER BY dbo_TBLSample.Station_ID, dbo_TBLSample.Start_Date;
```

The results are shown in Table 10. In this case, the query was constrained to a selected station and a date to limit the number of results. Other constraints such as year, water body name, 8-digit HUC code, monitoring organization, etc. can be used to retrieve data as needed.

Table 10 Available community indices for sample taken at Station 26 on 7/27/2000.

Station_ID	Start_Date	Indices_Group_Description	SumOfResult_Value
26	07/27/2000	Cyprinidae	6
26	07/27/2000	Insectivores	23
26	07/27/2000	Omnivores	40
26	07/27/2000	Simple lithophils	4
26	07/27/2000	Suckers	3
26	07/27/2000	Sunfish	17
26	07/27/2000	Top carnivores	23

Present Status and Future Work

Fox River watershed data, Minnesota PCA data, Lower Des Plaines and Lower Wisconsin data are available on the server SPRUCE18 as the MSAccess format and used in the project. Chemical water quality data for the Fox River watershed is available in STARED SQLserver. Biological data of Fox River watershed, Minnesota PCA data, and Lower Des Plaines, Lower Wisconsin data and STORET data for Great Miami River of Ohio are validated and reformatted into the proper structure for the purpose of taking into STARED. Maryland DNR data and Ohio EPA data are available and are being converted to the STARED format to be taken into SPRUCE18 server. The collection and compilation of biological endpoint data were found more difficult. Key decisions about what level of data to include, how to categorize species, how to deal with incomplete data sets and what should be the structure of the template for the data tables were needed to be made before the data were actually entered. The collection of all data and updating of the database will be completed by end of August 2005.

References

Adolphson, D.L., Arnold, T.L., Fitzpatrick, F.A., Harris, M.A., Richards, K.D., Scudder, B.C., and Stewart, J.S. 2002. *Physical, chemical, and biological methods and data from the urban land-use-gradient study, Des Plaines and Fox River Basins, Illinois, 1999-2001*: U.S. Geological Survey Open-File Report 01-459, 20 p

Genet, J. and Chirhart, J. 2004. *Development of a Macroinvertebrate Index of Biological Integrity (MIBI) for Rivers and Streams of the Upper Mississippi River Basin*. Minnesota Pollution Control Agency, Biological Monitoring Program. St Paul, Minnesota. 20p.

Illinois Environmental Protection Agency (IEPA). 2005. *River and Stream Monitoring Programs* (http://www.epa.state.il.us/water/surface-water/river-stream-mon.html) Accessed on 12 July, 2005.

Karr, J.R., Fausch, K.D., Angermeier, P.L., Yant, P.R. and Schlosser, I.J. 1986. *Assessing Biological Integrity in Running Waters: A Method And Its Rationale*. Illinois Natural History Survey Special Publication 5. 28p.

McConkey, S., Bartošová, A., Lin, L-S., Andrew, K., Machesky, M. and Jennings, C. 2004. *Fox River Watershed Investigation – Stratton Dam to the Illinois River: Water Quality Issues and Data Report to the Fox River Study Group, Inc*. Illinois State Water Survey. Contract Report 2004-06. March 2004 (http://ilrdss.sws.uiuc.edu/fox/fox_report_phase1.asp?ws=3) Accessed on 12 July, 2005.

Mercurio, G., Chaillou, J.C., and Roth, N.E. 1999. *Guide to Using 1995-1997 Maryland Biological Stream Survey Data.* Prepared by Versar, Inc., Columbia, MD, for Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. CWBP-MANTA-EA-99-5.

Minnesota Pollution Control Agency (MPCA). 2005. The Minnesota Pollution Control Agency. (http://www.pca.state.mn.us/water/index.html) Accessed on 12 July, 2005.

Roth, N. E., Southerland, M. T., Mercurio, G., Chaillou, J. C., Kazyak, P. F., Stranko, S. S., Prochaska, A., Heimbuch, D. G., and Seibel, J. C. 1999. *State of the Streams: 1995–1997 Maryland Biological Stream Survey Results*, CBWP-MANTA-EA-99-6, Maryland Department of Natural Resources, Monitoring and Non-tidal Assessment Division, Annapolis, Maryland.

Santucci, V. J. Jr. and Stephen R. G. 2003. *Fox River Fish Passage Feasibility Study: Final Report*. Submitted to: Illinois Department of Natural Resources. 368p.
(http://www.co.kane.il.us/kcstorm/dams/fishpssg/final.pdf) Accessed on 13 July, 2005.

United States Department of Agriculture (USDA), 2004: Integrated Taxonomic Information System (ITIS) (http://www.itis.usda.gov). Accessed on 20 July, 2004.

United States Environmental Protection Agency (USEPA). 2003. *STORET 2.0 Data Entry Module: User's Guide*. U.S. Environmental Protection Agency.

United States Environmental Protection Agency (USEPA). 2005. STORET home page. (http://www.epa.gov/storet) Accessed on 13 July, 2005.

United States Geological Survey (USGS). 2005a. National Water Information System home page. (http://waterdata.usgs.gov/nwis). Accessed on 13 July, 2005.

United States Geological Survey (USGS). 2005b. NAWQA Water Quality Data Warehouse. (http://water.usgs.gov/nawqa/data). Accessed on 13 July, 2005.

Wisconsin Department of Natural Resources (WDNR). 1999. *Land Cover of Wisconsin - User's Guide to WISCLAND Land Cover Data*. Wisconsin DNR. 1999 (http://www.dnr.state.wi.us/maps/gis/datalandcover.html#data). Accessed on 13 July, 2005.

APPENDIX

STARED Data Dictionary

Primary Table Descriptions

In STARED, tables are grouped into seven major categories according to the way they are described under the section, Database Structure. Among these groups, six groups are designated for the information on project, sample, station, results, parameter and taxa; the seventh table, *TBLIDLocations* is used to store the information of the original source of data provider.

In the following tables different data types are used to store the information.

Data Type	*Description*
bigint	An integer type. Storage size is 8 bytes
Char	A fixed-length non-Unicode character data type with length of *n* bytes. *n* must be a value from 1 through 8,000
Float	An approximate number data type for use with floating point numeric data
decimal	A numeric data type with fixed precision and scale
Int	A normal-size integer type. Storage size is 4 bytes
nvarchar	A variable-length Unicode character data type of *n* characters. *n* must be a value from 1 through 4,000. Storage size, in bytes, is two times the number of characters entered
smalldatetime	A date and time data type for representing date and time of day with the time portion only accurate to one minute
varchar	A variable-length non-Unicode character data with length of *n* bytes. *n* must be a value from 1 through 8,000. Storage size is the actual length in bytes of the data entered, not *n* bytes

Allowable length of a data type is also given in the brackets.

Tables

Data Source-Related Table:

TBLIDLocations		
Column Name	*Description*	*Data Type*
IDLoc	Code identifying data source/electronic file	nvarchar(1)
ID_Description	Code description	nvarchar(50)

Project-Related Tables:

TBLProjects_Programs		
Column Name	*Description*	*Data Type*
Project_Code	Unique code assigned to a project	nvarchar(50)
Program_Project	Name of the project for a particular monitoring effort	nvarchar(255)
Organization_ID	Code of organization conducting the project. Lookup - *TBLOrganization.*	int
Project_Study_Area	Description of the project study area.	nvarchar(50)
Project_Purpose	Description of the project, type of monitoring and intent, etc.	nvarchar(255)
Project_Start_Date	Starting date of the project (MMDDYYY)	smalldatetime
Project_End_Date	Ending date of the project (MMDDYYY)	smalldatetime
Contact_Name	Contact name - if appropriate/available	nvarchar(50)
Contact_Phone	Contact phone - if appropriate/available	nvarchar(12)

TBLProject_Grade		
Column Name	*Description*	*Data Type*
Project_Code	Unique code assigned to a project. Lookup – *TBLProjects_Programs*	nvarchar(50)
QAPPCode	Group of parameters for which a grade is assigned. Lookup – *TBLParameter_Group*	int
QAPP_Grade	Quality Grade assigned to the parameter group	int
CU_Grade	Comparability Grade assigned to the parameter group	int

TBLOrganization		
Column Name	*Description*	*Data Type*
Organization_ID	Unique number assigned to an organization	int
Organization_Short	Abbreviated name of the organization	nvarchar(50)
Organization_Name	Official name of the organization	nvarchar(255)
Organization_Type	Short narrative describing the organization	nvarchar(255)
Contact_Name	Contact name - if appropriate/available	nvarchar(50)
Contact_Phone	Contact phone - if appropriate/available	nvarchar(12)
Address	Organization address - if appropriate/available	nvarchar(50)
Zip	ZIP code. Lookup - *TBLZip*	int
Web_Site	Organization web site	nvarchar(50)

TBLZip		
Column Name	*Description*	*Data Type*
Zip	ZIP code	int
City	City	nvarchar(50)
State	State	nvarchar(2)

Station-Related Tables:

TBLStation_Information		
Column name	*Description*	*Data Type*
Station_ID	Unique number assigned to a station	int
Latitude	Latitude in decimal degrees	float(8)
Longitude	Longitude in decimal degrees	float(8)
Lat_Long_Accuracy	Code describing accuracy and origin of latitude and longitude. Lookup - *TBLLat_Long_Accuracy.*	nvarchar(2)
EPA_Station_Code	The station code used by the USEPA and the IEPA – if appropriate/available.	nvarchar(20)
USGS_Station_Code	The station code used by the USGS - if appropriate/available	int
Station_Code1	Reserved for other organization codes	nvarchar(50)
Station_Code2	Reserved for other organization codes	nvarchar(50)
Place_Name_Description	Descriptive information about the station site, such as bridge/road names, nearby communities, or features.	nvarchar(255)
TempRiver_Stream_Lake	Location information from Legacy STORET data, a merge of the three station_name fields. For reference only.	nvarchar(50)
Total Area	Drainage area of river/stream/lake at the station, in square miles, if available.	float(8)
Hydrologic_Unit_Code	Eight-digit code assigned by the USGS.	int
Station_Type	Code describing the feature where the station is, e.g. river/stream/lake etc. Lookup – *TBLStation_Type*	int
Water_Body_Name	Common name of the water feature where station is located.	nvarchar(50)
RF3_River_Reach	RF3 Reach Code of a river reach where station is located.	nvarchar(50)
NHD_24K_River_Reach	NHD Reach Code of a river reach where station is located.	nvarchar(50)
Ambnt	Indicates whether station is ambient	char(1)

TBLStation_Type		
Column Name	*Description*	*Data Type*
Station_Type	Unique number assigned to a station type. Links to *TBLStation_Information*	int
Primary_Type	Letter code describing the feature where the station is located, e.g. river/stream, lake, wetland, etc.	nvarchar(50)
Secondary_Type	Letter code, further describing the feature where the station is located, e.g., type of wetland, etc.	nvarchar(50)
Natural_Indicator_Type	Single character describing station site, Y= natural feature, N= artificial/manmade feature.	nvarchar(1)

TBLLandUsage		
Column Name	*Description*	*Data Type*
Station_ID	Unique number describing a station. Lookup - *TBLStation_Information*	int
Land_Use	Description (only available for select stations)	nvarchar(50)
Land_Use_Code	Code for Description	nvarchar(5)

TBLLat_Long_Accuracy		
Column Name	*Description*	*Data Type*
Lat_Long_Accuracy_Code	Unique code assigned to the accuracy of station location. Links to *TBLStation_Information*	nvarchar(2)
Accuracy_Description	Describes Accuracy of Latitude and Longitude	nvarchar(250)

TBLCFS_Daily		
Column Name	*Description*	*Data Type*
Station_ID	Unique number describing a station. Links to *TBLStation_Information*	int
Date	Date of measurement	smalldatetime
Discharge	Flow discharge amount (cfs)	decimal(9)
Source_Number	Source Number	int

TBLWatershed_Info		
Column Name	*Description*	*Data Type*
Watershed_ID	Unique number assigned to a watershed	int

Station_ID	Unique number describing a station. Lookup - *TBLStation_Information*	int
Drainage_area_sqmi	Drainage area in square miles	float(8)
Contributing_area_sqmi	Contributing area in square miles	float(8)
Slope_avg	Average slope	float(8)

TBLNHD_Reach		
Column Name	*Description*	*Data Type*
NHD_100K_River_Reach	NHD Reach Code, a unique identifier. Links to *TBLStation_Information*	nvarchar(50)
Waterbody_Name	Common name of the water feature	nvarchar(50)
Hydrologic_Unit_Code_8	Eight-digit code assigned by the USGS.	Int
Length_meters	Length in meters	Int
Elevation_Upstream	Elevation of the upstream point in feet	float(8)
Elevation_Downstream	Elevation of the downstream point in feet	float(8)
Slope_Reach	Slope	float(8)
StreamOrder_Strahler	Strahler stream order classification	Int

Sample-Related Tables:

TBLSample		
Column Name	*Description*	*Data Type*
UnqID	Unique code assigned to a sample	nvarchar(20)
IDLoc	Code identifying data source/electronic file. Lookup - TBLIDLocations	nvarchar(2)
Sample_Code	Original sample identification number. A sequential number is assigned if not provided with the original data.	bigint(8)
Project_Code	Unique code describing a a project. Lookup - *TBLProjects_Programs*	nvarchar(50)
Station_ID	Unique number describing a station. Lookup - *TBLStation_Information*	int
Start_Date	Date and time when field work/sample collection began, MMDDYYYY HH:MM	smalldatetime
Missing_Time	Indicates whether sampling time was missing in the original data.	varchar(1)
End_Date	Date and time when field work/sample collection ended, MMDDYYYY HH:MM	smalldatetime
Medium	Unique code describing the type of material sampled. Lookup - *TBLMedium*.	nvarchar(1)

Sample_Type	Unique code describing the type of the sample and the sampling method. Lookup - *TBLSample_Type*.	nvarchar(1)
Composite_Statistic_ Code	Unique code assigned to a statistic representation of the sample. Lookup - *TBLComposite_Statistic*.	nvarchar(1)
Replicate	Replicated sample information.	nvarchar(1)
Sample_Depth_Length	Depth at which the sample was taken or length of the sampled reach, in feet.	float(8)
Comment	Any comment. Also used to designate samples for which a replicate is available in *TBLReplicates*.	nvarchar(50)

TBLMedium		
Column Name	*Description*	*Data Type*
Medium	Unique code assigned to a type of the material sampled. Links to *TBLSample*	nvarchar(1)
Medium_Description	Description of the sampled material.	nvarchar(30)

TBLSample_Type		
Column Name	*Description*	*Data Type*
Sample_Type	Unique code describing the type of the sample and the sampling method (e.g., grab, continuous, spatial composite). Links to *TBLSample*	nvarchar(1)
Sample_Type_Description	Sample type description	nvarchar(150)

TBLComposite_Statistic		
Column Name	*Description*	*Data Type*
Composite_Statistic_Code	Unique code assigned to a statistic representation of the sample (e.g, average, minimum, maximum). This information is inherited primarily from Legacy STORET. Links to *TBLSample*	nvarchar(1)
Composite_Statistic_Name	Composite statistic name	nvarchar(3)
Composite_Statistic_ Description	Composite statistic description	nvarchar(255)

Results-Related Tables:

TBLResults		
Column Name	*Description*	*Data Type*
UnqID	Unique code describing a sample. Lookup - *TBLSample*	nvarchar(20)
Parameter_Code	Unique code describing the constituent measured. Lookup - *TBLParameter_Codes*.	int
Result_Value	Data value for a sample result. Result values must be numeric values.	float
Remark_Code	Unique code qualifying the result. Lookup - *TBLResults_Remarks*	nvarchar(2)
Grade	Used to flag questionable data.	nvarchar(2)
Comments	Comments	nvarchar(50)

TBLResults_Val_NonNumeric		
Column Name	*Description*	*Data Type*
UnqID	Unique code describing a sample. Lookup - *TBLSample*	nvarchar(20)
Parameter_Code	Unique code describing the constituent measured. Lookup - *TBLParameter_Codes*.	int
Result_Value	A single character code and definition qualifying the result. Lookup - *TBLResults_Remarks*	nvarchar(5)
Remark_Code	Unique code qualifying the result. Lookup - *TBLResults_Remarks*	nvarchar(2)
Grade	Used to flag questionable data.	nvarchar(2)
Comments	Comments	nvarchar(50)

TBLReplicates		
Column Name	*Description*	*Data Type*
UnqID	Unique code describing a sample. Lookup - *TBLSample*	nvarchar(20)
AltUnqID	Alternate unique code describing a sample for replicate results. Lookup - *TBLSample*	nvarchar(20)
Parameter_Code	Unique code describing the constituent measured. Lookup - *TBLParameter_Codes*	int
Result_Value	Data value for a sample result. Result values must be numeric values.	float
Remark_code	Unique code qualifying the result. Lookup - *TBLResults_Remarks*	nvarchar(2)
Grade	Used to flag questionable data.	nvarchar(2)
Sample_Code	Original sample identification number.	bigint
ALTSample_Code	Alternate sample identification number for	int

	replicate records	
IDLoc	Code identifying data source/electronic file. Lookup - TBLIDLocations	nvarchar(2)
AltIDLOC	Alternate IDLOC for Replicate Records	nvarchar(1)

TBLBiol_Taxa		
Column Name	*Description*	*Data Type*
UnqID	Unique code describing a sample. Lookup - *TBLSample*	nvarchar(20)
ITIS_Code	ITIS code describing the taxon. Lookup – *TBLITIS_Code*	int
Hybrid_ITIS_Code	ITIS code describing the secondary taxon of a hybrid. Lookup – *TBLITIS_Code*	int
Result_Value	Number of individuals caught.	float
Remark_Code	Unique code qualifying the result. Lookup - *TBLResults_Remarks*	nvarchar(2)
Grade	Used to flag questionable data.	nvarchar(2)
Hybrid	Used to flag hybrids	nvarchar(1)
Comments	Comments	nvarchar(50)

TBLResults_Remarks		
Column Name	*Description*	*Data Type*
Remark_Code	A unique code qualifying the result value. Links to *TBLResults, TBLReplicates, TBLBiol_Taxa, TBLResults_Val_NonNumeric*	nvarchar(2)
Remark_Description	Remark description	nvarchar(255)

Parameter-Related Tables:

TBLParameter_Codes		
Column name	*Description*	*Data Type*
Parameter_Code	Unique code assigned to a parameter	int
EPAGroup_Code	USEPA Group code. Lookup - *TBLEPA_Group_Code*	int
Reporting_Units	Unique code describing a reporting unit. Lookup - *TBLReporting_Units*	nvarchar(1)
Decimal_Point	Decimal point	int
Short_Name	Short description of the parameter	nvarchar(100)
Full_Name	Full description of the parameter	nvarchar(255)

TBLEPA_Group_Code		
Column Name	*Description*	*Data Type*
EPAGroup_Code	Group code from original Legacy data. Links to *TBLParameter_Code*	int
EPAGroup_Description	Description of code	nvarchar(255)

TBLReporting_Units		
Column Name	*Description*	*Data Type*
Reporting_Units	Unique code assigned to a reporting unit	nvarchar(1)
Reporting_Units_Description	Reporting units description	nvarchar(255)

TBLQAPPGroups		
Column Name	*Description*	*Data Type*
QAPPCode	Unique number describing a group of parameters. Lookup – *TBLQAPP_Group_Codes*	int
Parameter_Code	Unique code describing the constituent measured. Lookup - *TBLParameter_Codes*	int

TBLQAPP_Group_Codes		
Column Name	*Description*	*Data Type*
QAPPCode	Unique number assigned to a group of parameters	int
Media_Group	Unique number describing the sampled material. Lookup - *TblMedia_Group*	int
Parameter_Group	Unique number describing parameters. Lookup - *TblParameter_Group*	int

TBLParameter_Group		
Column name	*Description*	*Data Type*
Parameter_Group	Unique number assigned to a parameter group	int
First_Order_Parameter_Group	A general parameter group	nvarchar(50)
Second_Order_Parameter_Group	A specific parameter group	nvarchar(50)

TBLMedia_Group		
Column name	*Description*	*Data Type*
Media_Group	Unique number assigned to a sampled material. Links to *TblMedia_Group*	int
Media_Group_Description	Describes the sampled material	nvarchar(50)

Taxa-Related Tables:

TBLITIS_Code		
Column name	*Description*	*Data Type*
ITIS_Code	Unique number assigned to a taxon	int
Rank_ID	Unique number describing a specific level within the taxonomic hierarchy. Lookup – *TBLRank_Information*	int
Kingdom_ID	Unique number describing the highest level of the taxonomic hierarchy structure. Lookup - *TBLKingdom*	nvarchar(1)
Taxa_Name_1	The singular or first part of a scientifically accepted label for an occurrence of Taxonomic Units.	nvarchar(50)
Taxa_Name_2	The second part of a scientifically accepted label for a binomial/polynomial occurrence of Taxonomic Units.	nvarchar(50)
Taxa_Name_3	The third portion of a scientifically accepted label for a polynomial occurrence of Taxonomic Units.	nvarchar(50)
Common_Name	A name other than a scientific name that is commonly used to refer to a species or other taxon	nvarchar(50)
Parent_ITIS_Code	ITIS code for the taxon's parent . Lookup - *TBLITIS_Code*	int
Assessment_Group_Code	Unique code describing the assessment group. Lookup – *TBLAssessment_Group_Code*	nvarchar(1)
Native	Unique code specifying if the species is native	nvarchar(1)
Notes	Additional information	nvarchar(25)

TBLRank_Information		
Column name	*Description*	*Data Type*
Rank_ID	Unique number describing a specific level within the taxonomic hierarchy.	int
Kingdom_ID	Unique number describing the highest level of the taxonomic hierarchy structure. Lookup - *TBLKingdom*	nvarchar(1)
Rank_Description	Description of the rank	nvarchar(30)
Parent_Rank	Parent rank as defined by its kingdom of rules. Lookup – *TBLRank_Information*	int

TBLKingdom		
Column name	*Description*	*Data Type*
Kingdom_ID	Unique number describing the highest level of the taxonomic hierarchy structure	nvarchar(1)
Kingdom	Description of the kingdom	nvarchar(20)

TBLAssessment_Code		
Column name	*Description*	*Data Type*
Assessment_Code	Unique code assigned to a biological assessment group. Links to *TBLITIS_Code*	nvarchar(1)
Assessment_Code_Description	Description	char(50)

TBLNative_Code		
Column name	*Description*	*Data Type*
Native_Code	Unique code specifying if the species is native. Links to *TBLITIS_Code*	nvarchar(1)
Native_Code_Description	Description	nvarchar(40)

TBLTolerance_Categories		
Column name	*Description*	*Data Type*
ITIS_Code	Unique number describing a taxon. Lookup - *TBLITIS_Code*	int
Tol_Region_Code	Unique number describing a geographic region. Lookup – *TBLTol_Region_Code*	int
Tolerance_Code	Unique code describing the tolerance category. Lookup – *TBLTolerance_Code*	nvarchar(1)

TBLTol_Region_Code		
Column name	*Description*	*Data Type*
Tol_Region_Code	Unique number assigned to a geographic region	int
Tol_Region_Description	Description of the region	nvarchar(50)

TBLTolerance_Code		
Column name	*Description*	*Data Type*
Tolerance_Code	Unique code assigned to a tolerance category	nvarchar(1)
Tolerance_Code_Description	Description of the code	nvarchar(50)

TBLTolerance_Values		
Column name	*Description*	*Data Type*
ITIS_Code	Unique number describing a taxon. Lookup - *TBLITIS_Code*	int
Tol_Region_Code	Unique number describing a geographic region. Lookup – *TBLTol_Region_Code*	int
Tolerance_Value	Numerical value	float(8)

TBLIndices_Group		
Column name	*Description*	*Data Type*
ITIS_Code	Unique number describing a taxon. Lookup - *TBLITIS_Code*	int
Indices_Group_Code	Unique code describing the indices group	nvarchar(3)

TBLIndices_Group_Code		
Column name	*Description*	*Data Type*
Indices_Group_Code	Unique code assigned to a community indices group	nvarchar(3)
Indices_Group_Description	Description of the code	nvarchar(40)