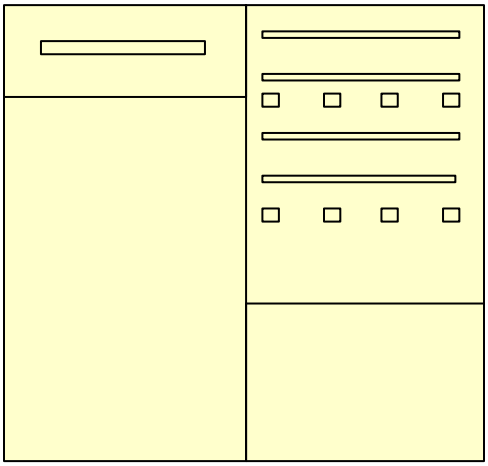
The KTOI Ecosystem Project Relational Database

A Report Prepared by Statistical Consulting Services[1] for KTOI Describing the Key Components and Specifications of the KTOI Relational Database



BPA PROJECT 1994-049-00
CONTRACT # 39133
Submitted September 2009

[1] Author: **Bahman Shafii**

Background Information

Data are the central focus of any research project. Their collection and analysis are crucial to meeting project goals, testing scientific hypotheses, and drawing relevant conclusions. Typical research projects often devote the majority of their resources to the collection, storage and analysis of data. Therefore, issues related to data quality should be of foremost concern. Data quality issues are even more important when conducting multifaceted studies involving several teams of researchers. Without the use of a standardized protocol, for example, independent data collection carried out by separate research efforts can lead to inconsistencies, confusion and errors throughout the larger project.

A *database management system* can be utilized to help avoid all of the aforementioned problems. The centralization of data into a common relational unit, i.e. a relational database, shifts the responsibility for data quality and maintenance from multiple individuals to a single database manager, thus allowing data quality issues to be assessed and corrected in a timely manner. The database system also provides an easy mechanism for standardizing data components, such as variable names and values uniformly across all segments of a project. This is particularly an important issue when data are collected on a number of biological/physical response and explanatory variables from various locations and times.

The database system can integrate all segments of a large study into one unit, while providing oversight and accessibility to the data collection process. The quality of all data collected is uniformly maintained and compatibility between research efforts ensured. While the physical database would exist in a central location, access will not be physically limited. Advanced database interfaces are created to operate over the internet utilizing a *Web-based relational database*, allowing project members to access their data from virtually anywhere. These interfaces provide users with the ability to upload, download, edit, and search data remotely, creating a dynamic system that is continually updated with the most recent information. At the same time, data are protected through user access restrictions, by implementing user profiles and password protected security. This accessibility could be set to any combination of read/write/edit abilities from an administrator capacity with full access to all data, to a highly restricted public access capability limited to general project information. Generation of customized summary reports and basic graphical routines could also be obtained through a Web-based interference. Using these types of features, users could produce summary tables, track trends of specified response variables over time or location, and compare results from various disciplines. Exploration of data in this manner can help users to better define and clarify their research goals and provide a means of integrating various aspects of a larger research project.

Kootenai Tribe of Idaho Fish and Wildlife Database

In December 2003, *Statistical Consulting Services* (SCS) was commissioned to create, customize, maintain, and operate a Web-based relational database for *Kootenai Tribe of Idaho* (KTOI). This included incorporation and operation related to all trophic level data and associated information for BPA Ecosystem, Operational Loss, and Kootenai Lake projects. In consultation with the project managers, a dedicated computer system, hosted by a reputable firm, *CI Host*, was acquired for this purpose. An initial Web page was then constructed and customized as per specific requirements of the clients**.** Bio-monitoring trophic level data generated by various projects were formatted, collated, and uploaded into the *designated components* of the relational database, as they became available. Exploratory summary and graphical routines were subsequently implemented for each of the specified components, as specified by the clients. More *sophisticated options*, such as data censoring, multi-year-trophic level plotting displays, dynamic maps, etc, were then incorporated on needs/available funding basis. User *profiles* were also created, and *security* was implemented at a level requested and specified by the project sponsors. The KTOI fish and wildlife database has been operational since March 2004.

The design of these WEB-based relational database systems allow specified users to access all data online as well as carry out some basic search, summary, censoring, and plotting routines. This design allows for independent access to each component as well as collective functionality for potential inter-trophic analyses. The Ecosystem database is designed around separate trophic level data components including algae, macroinvertebrates, fish, and water quality parameters, currently encompassing years 2001 to 2008. The current Kootenay Lake database includes components for water chemistry, phytoplankton, zooplankton, and mysid shrimp data covering years 2003 to 2008. Most recently, SCS has also incorporated data collected for the Fine-Scale Nutrient Addition Project, encompassing years 2005-2008 and consisting of water chemistry, chlorophyll, and periphyton data. The Operational Loss Relational database currently includes avian, terrestrial invertebrate, and site components, encompassing years 2002-2008.

The Ecosystem Relational Database

Technical Specifications

The Ecosystem relational database is housed on a dedicated Intel 3.0 GHz Pentium computer operated by *CI Host* of Los Angeles, USA. This computer is available 24 hrs/day, 365 days/year. Web services on the computer are provided by Apache Web Server software (ver. 1.1.37) running on RedHat Linux (ver. 7.1). The relational database was constructed using MySQL database software (ver. 4.1) and can be accessed via a web-based interface programmed in the PHP programming language (ver. 4.4) and HTML 4.0. The Ecosystem relational database may be accessed at:

http://www.scsnetw.com/ktoi/

Only authorized users, as specified by the respective project managers, are granted access to the system. Please contact Ecosystem Project manager Charlie Holderman directly for accessing information.

Current Data Availability and Functionality

Currently, the *Water* component houses data encompassing years 2002 through 2008. The 2009 data set requires further cleansing and data quality checks prior to incorporation. Typically, data on all variables are uploaded, however, only a pre-specified list of variables (as determined by the client) is available for list, search, summary and plotting routines. At this time a total of 2843 observations are recorded. Variables include *Site*, *Rep*, *NH4*, *NO2_3*, *SRP*, *TDP*, *TN*, *TP*, and *date*, representing site code, replication number, ammonia nitrogen, nitrate and nitrite nitrogen, soluble reactive phosphorus, total dissolved phosphorus, total nitrogen, total phosphorus, and calendar date, respectively. In addition to the above nutrient variables, the water component also includes data on metals for the specified period, i.e. Aluminum, Arsenic, Cadmium, Copper, Iron, Lead, Mercury, Selenium, and Zinc. Furthermore, information on river Temperature and Flow have also been incorporated into the water component of the relational database. The list, search, summary, and basic plotting routines are currently available for this component.

The *Algae* component currently includes data for years 2002 through 2007. The 2008 data set, while available, requires further data formatting and quality checks prior to incorporation. This component has been built as a three-tier component, consisting of information on chlorophyll, taxonomic groups, and ecological/biological metrics. There are a total of 1863, 38579, and 276 observations recorded, for each of the aforementioned sub-components, respectively. Variable list for the chlorophyll data include S*ite, Date, Replication, River Kilometers (RKM), Chlorophyll a, Chlorophyll a accrual rate, Total chlorophyll* , and *Total chlorophyll accrual rate.* Variable list for taxonomic groups, diatom and soft-body, includes *Site, Date, Taxa group,* and *Abundance*. There are numerous variables listed under the metrics sub-component, including various A*bundance, Richness, Dominance and Diversity* measures, and other physical and biological indices. The list, search, summary, and advanced plotting routines are currently available for the algae component of the relational database.

The *Macroinvertebrate* component of the relational database system currently houses data encompassing years 2002-2007. The 2008 data are being sorted, identified and processed at this time. Similar to the algae component, the macroinvertebrate component is a three-tiered one, consisting of aggregated, taxonomic and metrics data. There are a total of 1860, 300167, and 1815 observations recorded, for each of the aforementioned sub-components, respectively. Variable list for the aggregated macroinvertebrate data include S*ite, Date, Replication, River Kilometers (RKM), abundance* and *biomass*. Variable list for taxonomic data includes *Site, Date, Replication, Taxa group, Abundance,* and *Biomass*. As was the case for the algae data, variables listed under the metrics sub-component of the macroinvertebrates includes various *Abundance, Richness, Community composition, Functional group, and Diversity* measures, as well as other biotic indices.

The list, search, summary, and more advanced plotting routines are currently available for the macroinvertebrate component.

The *Fish* component of the relational database system currently includes data encompassing years 2002-2008. The 2009 data are being collected, sorted, and processed at the time of writing this report. Similar to the algae and macroinvertebrate components, the fish component of the relational database system is also three-tiered, consisting of aggregated, individual, and fish gut content information. Data on fish age, encompassing years 2002 through 2007 have also been collated and formatted for subsequent incorporation into a new fish sub-component of the relational database. There are a total of 12083, 6860, and 55271 observations recorded, for each of the aforementioned sub-components respectively, and 1945 observations available for the fish age data. Variable list for the aggregated fish data include S*ite, Date, Species, Sampling effort, Fish condition, Fish length, and Fish weight.* Variable list for the individual fish data includes the same information based on each individual fish species observation. Variable list for the fish gut content includes *Site, Date, Species* (note: predominantly Mountain Whitefish), *Content Taxa,* and corresponding, *Abundance*, and *Biomass*. The fish component of the relational database is rather complete and sophisticated, and will soon be augmented by additional valuable information such as fish length to age and relative weight. All system functionality, including list, search, summary, data censoring, and advanced plotting routines are currently available for this component.

The Ecosystem Project data availability information is summarized in Table 1.

Table 1. Ecosystem Project Data Availability Matrix.

	Year						
Trophic Level	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Water							
Algae							
Chlorophyll							
Taxonomic							
Metrics							
Macroinvertebrate							
Aggregated							
Taxonomic							
Metrics							
Fish							
Individual							
Aggregated							
Fish Gut							

Data available

Data collected, but not incorporated

Data not available

The *Fine-Scale* nutrient addition project was designed to monitor the potential water chemistry changes due to the addition of 3.0 .g·L-1 phosphorous, and assess algal production responses in the main-stem of Kootenai River. This project was designed to *supplement* the water chemistry and algal sampling of the KTOI bio-monitoring sites. The trophic level data available for the Fine-Scale nutrient addition project encompass years 2005 through 2008. The 2009 data are being collected and processed at this time. The lists of basic variables for the water quality and chlorophyll data are similar to those previously specified for the Ecosystem bio-monitoring project. The list of variables for the periphyton data includes S*ite, Date, Taxonomic groups, and Abundance*. Currently, only the list and search options are available for this component. Basic summary and plotting routines will be implemented in near future. The Fine-Scale Project data availability information is summarized in Table 2.

Table 2. Fine-Scale Project Data Availability Matrix.

	Year			
Trophic Level	**2005**	**2006**	**2007**	**2008**
Water				
Chlorophyll				
Periphyton				

Data available
Data collected, but not incorporated
Data not available

SCS started incorporating trophic level data generated by the Kootenai Lake Project into the Ecosystem relational database in the fall of 2007. The Kootenai Lake project is a component project of the larger Kootenai River Ecosystem Restoration Project. It involves mitigation of cultural oligotrophication to restore kokanee and other native fish populations, and restoration of habitats and supporting ecosystem functions in the South Arm of Kootenay Lake. This database includes various trophic components of limnological data from Kootenai Lake, including kokanee hydroacoustic surveys, and kokanee spawning escapement numbers from South Arm Kootenai Lake tributaries. Specifically, the current Kootenai Lake database includes components for water chemistry, phytoplankton, zooplankton, and mysid shrimp data covering years 2003 to 2008. The lists of variables for the various components of this relational database are as follows.

Water (Discrete-EPI): *site, date, field, lab, NH4, NO2+NO3, DIN, TDP, TP*, SI, *Chl-a,* and *Phaeophytin*

Water (Integrated): All the Discrete-EPI variables, plus *PH, Conductance, Turbidity, Alkalinity,* and *TOC.*

Water (Hypolimnion): Same as water-integrated

Discrete Phytoplankton: *Site, Date, Depth, Taxonomic Group, Abundance,* and *Bio-volume.*

Integrated Phytoplankton: *Data not available*

Zooplankton: *Site, Date, Taxonomic Group, Abundance, and Biomass.*
Mysid Shrimp: *Site, Year, Month, Abundance,* and *Biomass.*

All component trophic level data (encompassing years 2003-2008) have now been incorporated. Summary routines and specialized custom graphic displays are available for all the Kootenai Lake components.

The Kootenai Lake Project data availability information is summarized in Table 3.

Table 3. Kootenai Lake Project Data Availability Matrix.

	Year					
Trophic Level	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Chemistry						
Discrete						
Intergrated						
Hypolimnion						
Phytoplankton						
Discrete						
Integrated						
Zooplankton						
Mysid Shrimp						

Data available
Data collected, but not incorporated
Data not available

It is important to note that SCS has also incorporated limited data (water chemistry and phytoplankton) for the Arrow Lake Reservoir over selected years (2003-2005) under a separate sub-component of the Ecosystem relational database. The list of variables for the Arrow Lake water chemistry and phytoplankton data are exactly those of the Kootenai Lake. Only the list and search option has been implemented for this component, as per specific requirements of KTOI.

In addition to trophic level tabs providing linkage to the data and specified functionality on the KTOI main Web page, there are active tabs linked to site information, tools, downloads, and maps. The *site information* tab provides basic site description for all the bio-monitoring sites of the Ecosystem project. This tab could also provide a tabulated summary (matrix) of data availability for any and all requested trophic level data by year. The *tools* tab enables authorized users of the system to change/modify their passwords. The *download* tab provides access to data entry forms for all the aforementioned trophic levels. SCS has recently incorporated a working topographical map of the Ecosystem project's monitoring sites, under the *maps* tab, which is hot linked to the specified site description.

Future Data Incorporation and Enhancements

It is the intension of project sponsors to continue incorporating trophic level data into the relational database, i.e., water chemistry, algae, macroinvertebrate, and fish, as they are collected and become available. This will naturally include all other related data generated by the Fine-Scale and Kootenai Lake projects. Additional information on physical and ecological correlates will also be incorporated into the system. Possible future system enhancements include full text data descriptions and update for all incorporated components, implementation of data availability matrix for every component of the project, implementation of various mapping formats including topographic, GIS, etc, addition of data censoring option for all trophic level data, restructuring and enhancement of graphic capabilities (line plots, bar plots, pie charts), incorporation of multi-trophic/multi-year plotting routines, and implementation of more advanced security features.

Concluding Remarks

Collection, storage and management of high quality data are one of the most important aspects of conducting applied research. Researchers typically devote a majority of their allocated resources to data collection, cleansing, and analysis. Hence, issues related to data quality and integrity should be of outmost concern. Construction and maintenance of a centralized database management system continuously monitored and updated by a designated database manager, will elevate data quality assurance, and increase efficiency in dissemination of information. SCS created a Web-based relational database for KTOI to meet the aforementioned objectives, which has been operational since spring of 2004. All data components of this relational database, along with their associated features and functions have been updated over time. The Ecosystem Project relational database has provided a useful tool for the project managers as well as the sub-contractors. Continued efforts in upgrading and enhancing this system will ensure availability of quality data in real time, and validity of statistical analyses and interpretations for which such data are to be utilized.