

2006 summary of the Status, Trend and Monitoring database development for the Integrated Status and Effectiveness Monitoring Program  
Jeff Cowen

## **Background**

The Integrated Status and Effectiveness Monitoring Project (ISEMP) was initiated in 2003 with funding through the Bonneville Power Administration (BPA) in response to the need for status and trend and effectiveness monitoring called for by the 2000 Biological Opinion. Since 2004, ISEMP in the Wenatchee basin has focused on the design and implementation of a sampling regime and Status and Trend Monitoring program with 67 monitoring indicators. The monitoring project targets salmon and steelhead populations and habitat and is in collaboration with the Regional Technical Team (RTT), a shared interest group of over 10 agencies within the Upper Columbia basin. In fiscal year 2005, the scientific data management (SDM) team at the Northwest Fisheries Science Center, began a pilot project to develop a database repository and query interface to store field data collected through the ISEMP program in the Wenatchee subbasin. In fiscal year 2005, the SDM team focused on developing and populating a flexible and robust database structure to facilitate field data collected in 2004. In fiscal year 2006 SDM received \$200K in funding from BPA to continue working on the database structure and to develop a web enabled query interface to retrieve the data over the Internet. The purpose of this document is to highlight some of the major achievements and developments of the Status and Trend Monitoring database (STMD) for fiscal year 2006.

## **Data Exchange**

In order to promote data consistency and data exchange between data collectors and the STMD, the SDM team worked with NWFSC ISEMP data analyst to develop a series of MS Access Database templates. These templates can be used by the data collectors to help them manager their data and to help automate the data migration into the STMD. These templates provide a consistent data exchange and manage structure between data providers and the STMD repository. The SDM team is working on software to automate data migration between the templates and the database. Once these templates have been finalized they will become the standard mechanism for delivering data to the STMD and can be utilized throughout the region.

## **Database Developments**

Although the majority of the database design was completed in FY 05, several enhancements were added to the underlying structure. A journaling mechanism was developed so that a history of all changes to the data is maintained in the database. This structure will allow the database to track all changes to the data content so that when analysis is performed from the database those results can be reconstructed based on a state of the data at the time of the analysis.

A major component of this pilot effort has been the ability to track the protocol used in the field to collect a specific data element. In order to track these protocols at the data entity level the SDM team has been working closely with the Bureau of Reclamation to incorporate the BOR Protocol Manager data schema and protocol information into the STMD database structure. This year the Protocol Manager went through a database revision and therefore the STMD structure had to be modified to facilitate these updates. With these changes in place the STMD is able to store the data collection methodology implemented at every data collection event.

### **Database Query**

In order to meet the initial research needs of data users, the SDM team worked with NWFSC researchers and other ISEMP partners from the Washington State Department of Ecology (WDOE) to develop a set of query parameters for accessing the Water quality and Habitat data. It was determined that standard statistical information such as maximum, minimum, mean, and standard deviation, for Water Quality data would be required by the end user and that the user will be able to define the time span over which the data will be returned. It was also determined that the user will be able to return data over a running time span, so that a “running” average can be returned from the database over a user defined time span.

Over 80 data elements and calculations divided into 9 classifications were identified for the Habitat data. Many of the calculations were taken directly from the Washington state EMAP program and the WDOE. All of these derived calculations are fully documented insuring that the end user of this information will know exactly how the values were computed.

In order to provide access to both the water quality and habitat data, more than 50 materialized views or data queries were created in the database. These views automatically refresh once a week to reflect data changes. This mechanism insures that all of the data queries are documented and reproducible.

### **Web Interface**

While the data is the most critical aspect of the database, access to the data is an essential element. The field data is currently being collected under three broad subject areas: Habitat Characterization, Water Quality, and Fish Abundance. IESMP collaborators identified several statistical summarizations and over 80 calculations that were used as the basis to develop a web based query interface to access the data. To support this interface, the SDM team spent a considerable amount of time during FY 2006 in designing a software architecture that is maintainable and flexible as other data access requirements are identified and future releases are developed. This software code base was built using the industry standard Java J2EE platform and incorporates many Java frameworks such as Hibernate, Java Server Faces (JSF) and Spring.

The user interface that has been developed allows the end user to select specific data collection sites if their research scope is contained within a stream or subwatershed or they can choose to analyze all sites. An interactive map has been incorporated into the website so that the user can use the geographic location of a specific site and the hydrography of the basin to select additional sites upstream of a selected site. The user can select sites one at a time by selecting them from the map, by using the stream network to select points upstream of a selected site, and by selecting sites by data category. Once the sites have been determined, the user can select the category of data that they are interested in (water quality or habitat). After designating the data category, the user will be presented with various options on how to further refine their query to extract just the data that they are interested in.

Water quality data is collected from in stream data collection gauges which record information every hour. When querying the data, the user has the option of specifying a timeframe down to the day over which the water quality data will be accessed. The user has the option of retrieving the data in a raw data format or having the database perform a common set of summaries on the data including maximum, minimum, mean, and standard deviation. These statistics can be calculated as a daily rate or can be computed as a running calculation over a user specified range, so that computing the seven day running average over a two month period is possible. Once the timeframe has been established, the user can select specific data elements from a list of attributes which includes: acidity, average flow, conductivity, dissolved oxygen, temperature, and turbidity.

Habitat data sites are visited in most cases once a year and in subsequent years these sites may or may not be visited again as determined in the sampling schedule. Therefore habitat data can be queried by year only, although the exact date when the data was collected will be reported in the results of the query. Habitat data has been categorized into 9 classifications which were determined by ISEMP data analyst at the NWFSC. These categories include: Channel Morphology, Stream Profile, Substrate Size, Substrate Composition, Fish Cover, Large Woody Debris, Canopy Cover, Riparian Vegetation, and Human Disturbance. Within each classification the user can download the raw data or they can choose to have the data returned as any of more than 80 data summaries which are based on specific calculations that were established by EMAP and the Washington state Department of Ecology. All of these data calculations will be fully documented so that when the data is returned in this format a complete explanation of the computed result is available.

A software testing plan has been developed based on the user access needs to ensure that the web interface has been thoroughly tested before it is released to the user community.

## **GIS Development**

All data collection sites contain GPS coordinates to capture the spatial location of the site. Habitat data is collected as a series of stream transects, 5 upstream and 5 downstream transects are established from a central transect or X-Site. This central

transect location is used to represent the collection site. Water quality data is collected at fixed gauge locations in the stream, so the coordinates of this fixed location is stored. In either case, these coordinates were used to place the data collection site locations on a hydrologic network within a GIS. The hydrologic network was built using the 1:100K National Hydrologic Datasets (NHD) developed by the USGS and the ArcHydro toolset developed by ESRI and the University of Texas at Austin. A script was developed to run in the GIS software that stores the upstream sites for every site in a database table. This allows these hydrologic relationships to be accessed in the database without having to dynamically go through the GIS software.

### **Collaboration Tools**

In FY 06 the Wenatchee RTT identified the need to enhance the ability to share information among members. The Oracle Collaboration Suite (OCS) is a web based communication and information sharing tool that will allow user to post documents and data to a central repository and provide collaborators with access to this information. The OCS also provides several secure web based communication tools that allow users to chat and share desktop environments through a secure Internet channel. SDM team members have been working to deploy this application in a development environment and plan to make this collaboration suite available to the RTT in first quarter of FY 07.