System Integration to Support Effectiveness Monitoring of Salmonid Recovery Projects

Background

The goal of this pilot project is to establish a baseline database to manage data in support of research, monitoring and effectiveness of salmonid restoration projects in the Columbia Basin. In an effort to achieve functioning aquatic ecosystems and sustainable natural fish populations in the Upper Columbia River basin, NOAA research scientists at the Northwest Fisheries Science Center (NWFSC) have identified the need to develop a data management system to maintain biological, environmental and habitat characteristic information pertaining to aquatic natural resource management and restoration. This data system must be available to wide range of biologists, natural resource managers and researchers and must be accessible over the Internet through an easy-to-use query interface. To achieve this goal a development team at the NWFSC worked closely with database designers and GIS staff from the Bureau of Reclamation in Boise, Idaho to develop a pilot system that integrates a relational database management system with a GIS for the Wenatchee Subbasin.

The Wenatchee Subbasin provides spawning and rearing habitat for Endangered Species Act-listed Spring Chinook and Steelhead. The watershed originates in the high mountainous regions of the Cascade Crest and drains a portion of the east Cascade Mountains in north-central Washington. The mainstem of the Wenatchee River flows generally in a southeasterly direction, emptying into the Columbia River. The watershed encompasses approximately 1,327 square miles, with 230 miles of major streams and rivers.

Figure 1 Wenatchee Subbasin

In order to develop this pilot system, it was determined that this effort would focus on field data collected during the 2004 field season and would not incorporate any data previously been collected, although the database design has been developed to incorporate this data. It was also stated in the contracts that these data be collected following established data collection protocols and methods and the database must be able to track the protocol used to collect each piece of data. The NWFSC has received data collected by the U.S. Forest Service, Washington Department of Ecology, Washington Department of Fish and Wildlife, Chelan County Conservation District, as well as a number of private contractors. A set of 300 randomly located data collection sites was produced by the US Environmental Protection Agency (EPA) and it was decided that 50 of these sites would visited in 2004. In subsequent years 25 of these sites will be revisited and an additional 25 sites will be visited for the first time. A monument marker was established on land based on the coordinates provided by EPA for each site visited and a GPS coordinate was captured for where the work done in the stream took place.

Database Design and Data Collection

To store and retrieve the quantity and variety of information that is recorded at each field location a tabular database system comprised of over forty tables is maintained in an Oracle relational database. Every record in the database is associated with a field monitoring site whose location is maintained in the GIS. To maintain the relationship between the tabular and spatial data, a monitoring point table in the tabular database stores the unique identifier of the GIS point, which is stored in an SDE layer in another database.

At 50 monitoring reaches of these sites a series of ten transects were collected to assess the suitability for salmonid rearing habitat. In-stream habitat characteristics such as the amounts of large woody debris, sediment content and canopy type found along a segment of the stream were collected. In order to be able to reconstruct these transects in a GIS, a GPS coordinate was taken at the center of the transect and all other locations along the transect are stored as an offset from this location. All of this spatial information is stored in the tabular database and the GPS coordinates are used to generate a representative location in the GIS. Summary information for each transect is also derived and stored for fast retrieval from the information along the complete transect. The length of the transect is determined by accepted protocols and is based on the wetted width of the stream at the point of the center of the transect. This baseline information was collected for each transect during the 2004 field season and a subset of these sites will be visited next year, while information for new monitoring sites will be added each year.

Biological information for salmonid juvenile out-migration and macro invertebrates was collected at five locations in the watershed. These data were collected daily for up to 9 months (period of time that the rive is ice-free) and the information is stored in the tabular database, while the GPS coordinate for where in the stream the smolt trap or other data collection device was placed. In addition, at these five sites, a water monitoring station was set up to capture dissolved oxygen, pH level, and temperature every hour. In

order to this continuous temporal data a time-series table is maintained in the tabular database and a point location is maintained in the GIS.

ArcHydro

The ability to maintain the linkage between the spatial data and tabular data and the ability to query the data by spatial location and relative stream network location is critical to the utility of this system. Researchers want to be able to query monitoring points based on their relative in-stream position to other monitoring points in the system. The 1:100,000 scale National Hydrography Dataset that is developed by the USGS was used as the stream network on which the monitoring points are located. This dataset contains attributes which indicate the flow direction of each arc. The ArcHydro extension was used to establish the network flow. Once the stream network and flow direction was defined the monitoring point locations were snapped to this network and connectivity between the points was validated using Network Analyst tools. ArcHydro tools were then used to generate the unique identifiers (Hydroid) that are used to establish the relationship between the spatial locations and the tabular data. ArcHydro insures that all features within a network have a unique identifier within the network. These Hydroids are then used to maintain the relationship between the spatial locations and the tabular data.

To be able to query locations based on their upstream and downstream locations relative to other monitoring points, a custom ArcObjects script was developed. This script implements the ArcObjects iTraceSolver object's FindFlowElements method to determine all of the upstream monitoring points (junctions) from all other monitoring points. The results of this operation are stored in an SDE table that is linked to the tabular database. This table provides the linkage between the spatial locations and the tabular data and provides an efficient query mechanism of the spatial data in an SQL environment.

HYDROID	UPSTREAMHYDROID
1	2
1	4
1	19
1	20
1	31
3	2
3	4
3	19
3	20
3	31

Figure 2 Table produced from ArcObjects code showing associated upstream hydroids.

Web Interface

One of the most critical aspects of this pilot project was the ability to provide access to the data once it had been collected. Users of this data are distributed throughout the region, so a web-based platform was the obvious choice for distribution. This web

interface had to be intuitive to use and flexible enough to provide users with the query tools that would allow them to easily select and download the data that they need. This system needed to be available through all web browsers and should not require any additional plug-ins or software. To meet these needs a Java J2EE web query interface that is linked to an ArcIMS internet mapping client through the Hydroid lookup table was developed.

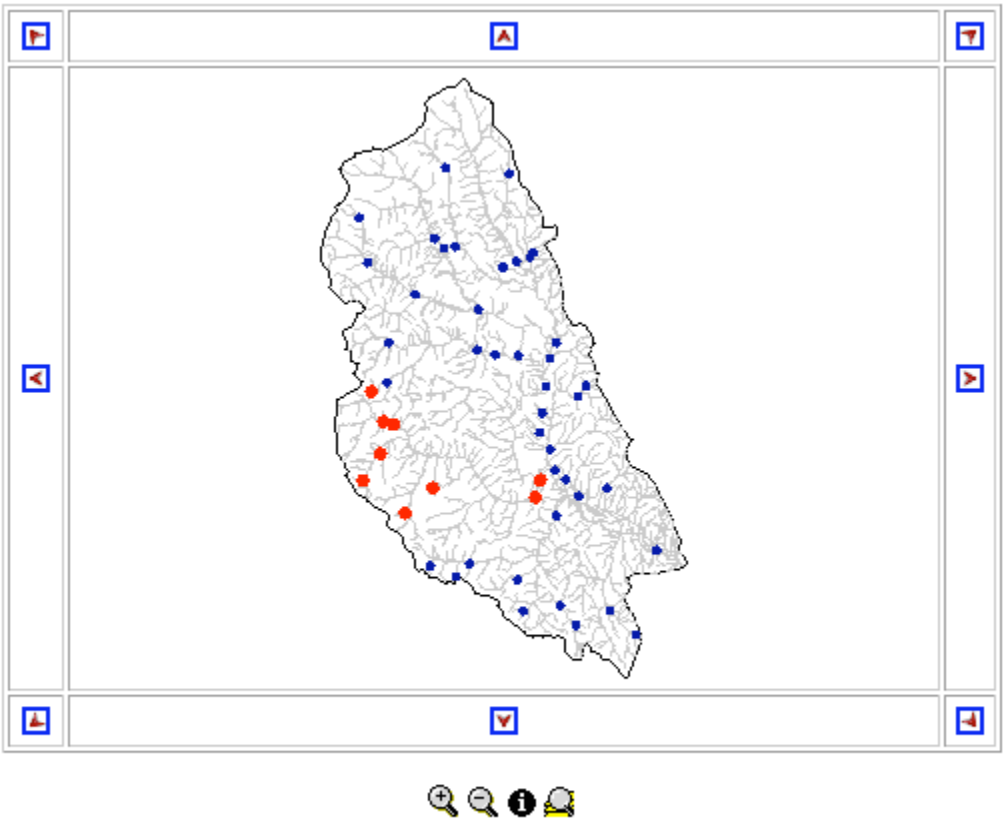
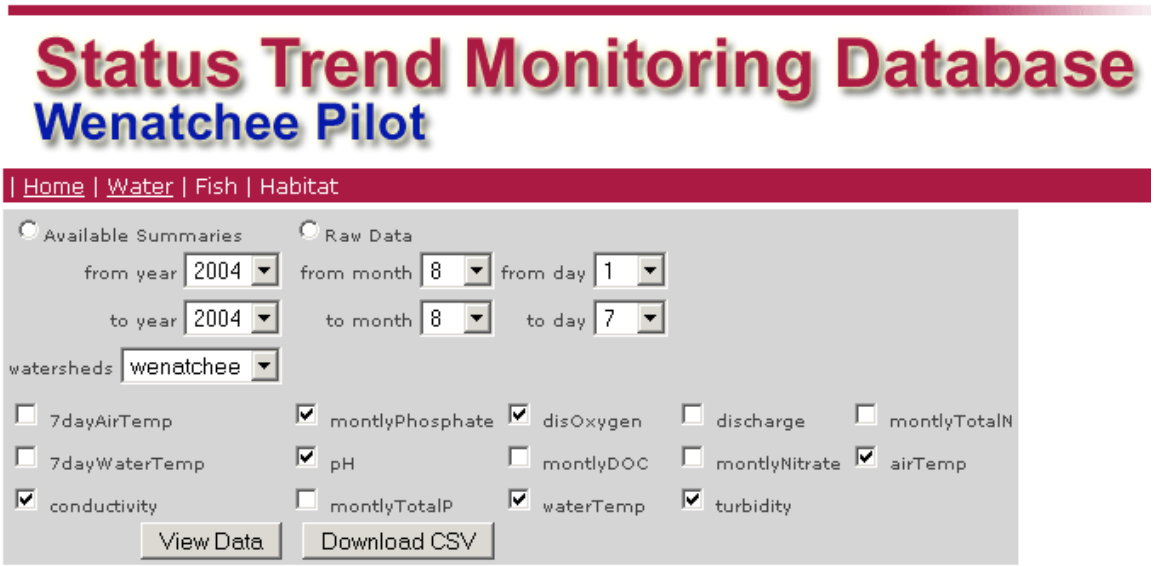


Figure 3 ArcIMS interface showing selected points upstream.

Through this interface users are able to interactively select a monitoring point from the map. The Hydroid of this monitoring point is passed to the database and related points upstream or downstream are determined through the Hydroid lookup table. These Hydroids are passed back to the IMS client to be displayed on the map. These Hydroids are also passed to the database query interface so that users are able to further summarize and filter the data based on the monitoring sites returned from the spatial query. For example researchers might be interested in the seven day running average water temperature for the two weeks prior to a smolt counting event. In this case the spatial filter can determine all of the water quality stations upstream of the smolt counting location (figure 3) and the query tools can narrow those results to calculate the seven day running average (figure 4). Since the network relationships are stored in a related table no specialized spatial tools had to be developed on the server side and the spatial filter can be implemented through a very efficient SQL query.



pH : Sun Aug 01 05:00:00 PDT 2004 - Sat Aug 07 12:00:00 PDT 2004

Min	Mean	StdDev	Date	Count	Max
7.84	7.9983335	0.14971470453373373	2004-08-01 00:00:00.0	24	8.23
7.85	8.0225	0.15629542817870656	2004-08-02 00:00:00.0	24	8.24
7.84	8.02125	0.16614065544382672	2004-08-03 00:00:00.0	24	8.27
7.84	8.032917	0.16361352040088561	2004-08-04 00:00:00.0	24	8.3
7.86	8.032917	0.1582370034959027	2004-08-05 00:00:00.0	24	8.28
7.89	8.01875	0.109596592298658	2004-08-06 00:00:00.0	24	8.17
7.94	7.94	0.0	2004-08-07 00:00:00.0	1	7.94

Figure 4 Query interface used to produce summary statistics on selected locations.

Conclusion

The first version of this system is scheduled to be released to participating research scientist and biologists in the Wenatchee Subbasin in July of 2005 and field data collected during the 2005 field season will be loaded into the database in early 2006. As more data is added to the system over time it should be possible for researchers to evaluate the effectiveness of restoration and monitoring efforts in the Wenatchee Subbasin. Once this system has been reviewed and tested by other data collectors and researchers, it will be determined if this system can be combined with other data collection efforts in the Columbia River Basin. At this point it is thought that this system will be able to be expanded to store previously collected data and with protocols in place this system may potentially be able to store all related data collected in the future.