

# Data Management Issues for Consideration During the NWPPC Fish and Wildlife Program Amendment Process

by:

The StreamNet Steering Committee

## Introduction

In recent years, the Fish and Wildlife Program (FWP) has supported several projects that acquire regionally valuable fish and wildlife related data and other information, consolidate them into a readily useable form, and make them available to managers, researchers, decision makers and the general public interested in Columbia Basin fish and wildlife issues. Those existing projects help to overcome a number of problems with acquiring and using data regionally, and they have also gained understanding of a number of remaining problems and issues that could be improved or rectified through amendments to the Fish and Wildlife Program. As one of those projects, StreamNet has gained experience with the strengths and weaknesses of fisheries data dissemination within the basin, and would like to identify a number of issues that could be addressed by amendments to the program.

### Issue 1: Need for regionally standardized data collection and data management

#### Issue Definition:

Fish, wildlife and habitat management in the Columbia Basin is the responsibility of a large number of different agencies and organizations. These include state and federal fish and wildlife, land management, water management and environmental agencies; tribal fish, wildlife and land management programs; various councils and commissions; private land managers; and a growing number of watershed groups. Each agency/entity is independent and has a unique area of responsibility and authority, geographic and/or legal.

At the present time, there are many kinds of data that are being collected across the basin by multiple agencies. However, in many cases there is little standardization of field methodology used to collect the data, or standardization of data definitions and formats for managing and sharing the data. Thus, seemingly similar data are not always comparable between (and sometimes within) agencies, making it difficult to accumulate data from multiple sources for use in performing wider scale analysis.

Use of different field methodologies can yield results that are different as much or more due to the specific methodology than the actual conditions, and comparisons can be made only if those differences are well understood and accounted for in the analysis. It also increases the risk of people attempting to use data for unsuitable purposes.

Non-standardized data definitions and formats make accurate analysis impossible unless units (such as count per mile v. count per transect, lb v. kg, etc.) and formats are standardized and equivalent. This situation was one reason for initiating the current data management programs, which either acquire data in a standard required format from individual sources, or acquire data and then convert them to a standardized format.

There is presently little incentive for agencies to go to the effort to standardize data collection and data management. Unless part of a regional effort, data are often collected primarily to meet a need related to the agency's particular mission or area of responsibility. Once that need is met, it may not be high priority to make the data available to others or to ensure that it is compatible with similar data collected by other agencies outside their area of responsibility. This leads to three possibilities: 1. A need to provide incentives and encouragement for agencies to work together on standardization; 2. A need for programs to acquire data from multiple sources, standardize them to the degree possible and make them available for use on a regional scale (like the existing data management projects); or, 3. Leave it to each potential data user in the basin to acquire data from multiple sources and be on their own to understand and correct the lack of standardization. Unless the preferred choice is option 3, it will be necessary to continue using the existing data management programs in option 2 while efforts are underway to encourage the agencies to agree on the need for standardization and undertake efforts to achieve it. The existing programs will also need to continue to maintain and make their archived data available to users.

#### Options for Resolution:

Our experience has shown that many people in need of basin wide data, while often qualified in technical fisheries issues, are not experienced in data management. Many have difficulty acquiring data from centralized computerized sources, let alone from multiple sources with inconsistent or incompatible data. Therefore, the primary options are continued support of data management programs to acquire and standardize data from disparate sources, coupled with active efforts to encourage agencies to work together to increase the standardization of common data.

The existing data management programs effectively deal with the data standardization issue for specific types of data. However, there are other kinds of data that are not yet included and developing standards among the agencies would simplify acquiring and posting data on a regional basis. Plus, the data management programs have no means of dealing with differences in field data collection methodology.

Since the many agencies in the basin are independent, it is not possible to mandate standardization. However, the Fish and Wildlife Program, as a regional program that works with all entities in the basin, is in a position to encourage the agencies to work together to address this issue. This could be done through sponsorship of workshops to reach regional agreement on how to deal with collection and management of specific types of data. In addition, the FWP funds a significant amount of fish and wildlife related work in the basin and therefore is in a position to require standardization for specific

types of data collected by projects receiving funding from the program. Specific formats could be required for core data types, similar to the data exchange formats used for exchanging data with the StreamNet database. Developing such exchange formats should be done in a collaborative process with the agencies, and the data management projects would be able to provide significant assistance to the process.

Developing standards for specific data types for projects funded by the FWP would go a long way toward developing true regional standards. Once standards are in place for program funded projects, it would be relatively easy for agencies to decide to adopt them for more widespread use within their programs, although doing so would be their choice.

## Issue 2: Need for regional priorities for data collection and availability

### Issue Definition:

Some kinds of basic fisheries and fish habitat data are widely recognized as priorities for collection and dissemination. Obvious priorities include hatchery production, various escapement measures, species distribution, dam passage counts and timing, harvest estimates, basin wide tagging information, water temperatures, and stream discharge. These kinds of data and more are already being compiled and distributed by the existing data management projects. However, there is presently no specific mechanism for establishing priorities for what data need to be collected in the field or what existing information needs to be acquired to address current priority issues in the basin.

In some cases, the lack of a priority setting process leads to expectations for specific data from the data management projects that were never conveyed to the projects. In other cases, needed data have not been collected in the field, or they have not been collected in a way that is suitable for addressing priority issues. For example, it was decided to develop assessments and management plans based on subbasins within provinces, but several types of fish information (for example, fish returns) are not collected by the management agencies on a by-subbasin basis.

In regards to the StreamNet project specifically, there is no mechanism for organized input into what the current priorities for data acquisition are. Therefore, it is difficult to decide what new kinds of data need to be added to the StreamNet database, or if other changes are needed. One potential solution would be for the StreamNet Steering Committee to annually convene a meeting and invite the primary fish and wildlife management agencies to participate in setting new data priorities. However, experience suggests that simply asking what data are needed does not necessarily lead to establishment of priorities, since there are many needs for data among agencies with different missions, and data are needed at all scales, from local to basin wide. Furthermore, the StreamNet Steering Committee is not in a position to establish basin wide data priorities for the many entities in the basin.

The need for specific data falls out from the questions or issues being addressed. What is needed, then, is a means to establish regional priorities for the specific issues or questions that are critical to address. Once the priority issues are identified, the data that are needed to address them become the priority data needs. The next step would be to determine whether the needed data already exist, or whether new data collection efforts need to be initiated. If new data collection efforts are needed, these could be used by the FWP to establish funding priorities. If the data exist, it should be straightforward to determine where they are, what formats they are in, and what is the simplest and most effective way to obtain them. Depending on the exact situation, there may or may not be a need for one of the data management projects to undertake acquisition and dissemination of the data.

#### Options for Resolution:

Because of its basin wide scope, the FWP is in a good position to call for basin wide establishment of priorities for issues and problems critically needing attention. Actual conduct of such an effort might be a logical role for the Columbia Basin Fish and Wildlife Authority, which encompasses the primary management agencies in the basin. A workshop including CBFWA members, other agencies, academics, and data managers might offer an effective means for establishing basin wide priorities for issues and problems needing immediate attention. Determining priority data needs would then fall out from the priority issues, and would provide guidance for agencies that collect the data and projects that acquire and disseminate the data.

A second option would be for one of the data management projects to convene a meeting to establish priorities for data needs. This could become an annual effort to maintain focus on providing data as needs evolve and change. It would provide needed guidance to the data management projects. Unless specifically tasked by the FWP, however, this approach would have lesser standing in the region for establishing priorities for current basin wide issues and problems.

### Issue 3: Need to use new technology to improve data management effectiveness

#### Issue Definition:

Technological advances are improving means of communicating, providing new opportunities for management of fish and wildlife information. To take full advantage of such technology, however, will require changes in how the entities in the basin handle their data collection and data management activities.

Geographic Information Systems provide enhanced means of displaying and conveying information and offer new means of analyzing data on a geographic basis. To utilize this technology, however, requires that all data be collected with geospatial references. Much traditional fisheries data has been collected with only general descriptive information on location; new efforts to collect specific location coordinates are widely needed.

The Internet is becoming a primary tool for disseminating information. Access to computers and the Internet is becoming widespread and is nearly complete within fish and wildlife related agencies and organizations, making this a nearly universal means for accessing data and information. The existing data management projects are already using the Internet to disseminate their data.

A concept that promises to improve the ability to locate and acquire information over the Internet is the use of specially designed search engines to locate metadata (information about the content, source, formats, and other characteristics of a data set) for data that are posted on the Web. As more entities are persuaded to post their data and to create and post metadata, this approach will make it easier to locate data on the Web, assess whether they are suitable for the need, and obtain the data. Data originators retain responsibility for maintaining and updating their data. As a tool for disseminating fish and wildlife related information, however, there are a number of limitations that need to be addressed if we plan to rely on this as a primary means of distributing information.

The Internet is anonymous. Data are transferred without direct contact between data provider and user. That makes it difficult to obtain feedback on how well the data met the needs of the user, or whether additional data or services are needed. The data are provided to all who access it, regardless of their qualifications and ability to analyze or utilize the data accurately and appropriately. While many qualified data users are becoming more adept at handling data over the Internet, many others are unfamiliar and untrained, and have a difficult time obtaining the data they need, even from existing sites intended to provide data quickly and easily. Our experience has been that many users need detailed help with how to locate, evaluate and download data over the Internet, suggesting that just posting data on the web will not be sufficient to meet all users' needs.

Fish and wildlife agencies currently do not post much of their data on the Internet. As mentioned in Issue 1, data are often collected primarily for the needs of the agency, and there are few incentives to take the time and effort to post and maintain the data once the primary need has been met. Also, those collecting data may be reluctant to openly disseminate data until after they have been fully analyzed and published. This general unavailability of fisheries data is one reason the data management projects were initiated. It will be necessary to encourage agencies to provide their data over the Internet, and to create and post metadata, before Web based search capabilities can be relied on to locate and obtain the information. Until that happens, the data management projects will need to continue to acquire, organize, standardize and provide data on behalf of the agencies.

Also as mentioned above, the present lack of standardized data collection methodologies, data definitions and formats between agencies make it difficult for data users to utilize data once they are posted on the Internet. Each entity acquiring data for a regional scale analysis will be on its own to figure out how to combine disparate data from multiple sources correctly for the intended analysis. Independent data sets will be developed by many data users, leading to duplication of effort. This increases the risk of similar analyses being conducted on data sets that were constructed differently leading to different results, adding confusion to the already complex decision making in the basin.

## Options for Resolution:

It seems apparent that there is a need to encourage the agencies collecting data in the Columbia Basin to not only standardize their data collection methodology and data formats for key kinds of data (Issue 1) but to also post and maintain their data, along with metadata, on the Internet. As with the issue of standardization, however, the agencies are independent and can not be mandated to do so. But, the FWP could play a role as a catalyst to encourage the agencies, possibly through CBFWA, to increase emphasis on development of metadata and posting data sets on the web.

More directly, the FWP could require that projects funded through the program post their data and metadata on the Internet, either through their own agency or through one of the regional data management projects. If data exchange formats have been developed as suggested in addressing Issue 1, those formats should be required for posting the data on the Web to provide consistency between different agencies collecting similar data.

Another approach would be for the FWP to require that projects funded through the program make their data available electronically, along with metadata and/or a data dictionary, for placement in a data archive. This would be similar to the current requirement by BPA that annual and final reports are to be provided in electronic format. The archive could post the data on the Internet if requested and funded to do so, or simply post the metadata. Such an archive could be developed within an agency or with one or more of the regional data management programs. Advantages to using the data management projects are the already existing servers and experience.

### Issue 4: Need to integrate data management and data analysis

#### Issue Definition:

Data analysis and data management are two separate but related activities. Since database management is a relatively new technical field, it is understandable that its capabilities and requirements may not always be fully understood. A common misunderstanding is that database management is equivalent to data analysis. However, while good database management can make analysis much easier and efficient, it is not an analysis process. Rather, it is a means of organizing, storing and retrieving data that can then lead to efficient use of the information.

Computer database programs have become quite powerful and are effective tools, but also have specific requirements for data that are not always understood by novice users. It is important for those conducting large data analysis projects to include database specialists early in project development to assist with correct acquisition and storage of the required data. Such assistance can identify problems with data early and can speed analysis. Having database management specialists working with an analysis project will also simplify the process of making both the raw and analyzed data available for others to use, either directly on the Internet or through one of the data management projects.

### Options for Resolution:

The Fish and Wildlife Program could emphasize the value of incorporating database management expertise in project proposals during the review and approval process. Proposal formats could be amended to require information on how project proponents intend to store and manage their data and how they intend to make their data available to others. Projects would be free to include their own expertise in database management, or could address the issue through collaboration with one of the regional data management programs. This information could then be included in consideration of funding decisions.

### Issue 5: Need for an electronic library/archive

#### Issue definition:

The volume of scientific information, collected by the various organizations and researchers in the Basin, needs to be collected and organized in a central location. The StreamNet Library already provides this service and should continue the effort. The availability of the Library to provide access to source documents related to scientific research relevant to fish and wildlife in the Basin needs to be more widely known.

Information has mainly been distributed in paper format, limiting the ability of researchers, managers and decision makers to access these documents in a timely fashion. Given the size of the Basin and the distance to the Library from many field locations, there is a need to increase the amount of materials accessible electronically. Historic reports and other significant documents need to be converted to electronic format for preservation and distribution purposes.

#### Options for resolution:

While the Library will acquire, archive and distribute reports from FWP funded projects and other sources, there needs to be a Basin-wide effort to convert historically significant documents to electronic format. FWP funded projects are already required to submit annual and final reports in electronic format. The FWP should expedite the conversion, beginning with the rare, historic documents of regional significance, by including such activities in the amended program.

### Summary of Recommendations

The following summary of recommendations is intended to offer specific suggestions on ways the Fish and Wildlife Program can amend its program to address the issues discussed above. The large number of entities collecting data in the Columbia Basin, and the large number of entities that need to use that data, make it important to develop effective programs that assure data are captured and made widely available in the easiest and most effective manner possible. These suggestions are offered with that end in mind.

#### Issue 1: Need for regionally standardized data collection and data management

- The FWP should encourage Columbia Basin agencies to work together to develop mutually agreed on standards for data collection methodology and data definitions for specific types of priority fisheries and wildlife data. CBFWA could take a lead role in this effort, and a series of workshops might be a productive approach.
- The FWP should require projects funded under the Program to provide specifically defined kinds of fisheries and wildlife related data to the Program in regionally standardized formats. Formats should be developed cooperatively among the agencies through workshops, and assistance with developing standardized formats can be obtained from the regional data management projects.
- The FWP should emphasize the continued value of the regional data management programs to acquire, standardize and disseminate data. The role of these projects should be reevaluated after regional standardization is achieved

#### Issue 2: Need for regional priorities for data collection and availability

- The FWP should institute a process to work with entities in the basin to annually determine the top priority issues and problems that require attention. These priority issues would then be used to guide FWP funding decisions and to define the priority data needs to guide regional data management projects in what data should be acquired and made available.

#### Issue 3: Need to use new technology to improve data management effectiveness

- The FWP should require projects funded by the Program to provide their raw and analyzed data, with metadata, in electronic format and have it posted on the Internet. Posting could be through the project's agency, a cooperating agency, or one of the regional data management projects.
- For data where standardized data formats have been established, the FWP should require data to be posted in those formats.
- The FWP should encourage fish and wildlife agencies in the basin to begin posting and maintaining core data types on the web, either through the agency or through the regional data management projects. This should be done collaboratively through CBFWA.

#### Issue 4: Need to integrate data management and data analysis

- The FWP should encourage projects funded through the FWP to involve database management as an integral part of project development by requiring the issue to be addressed in project proposals.

#### Issue 5: Need for an electronic library/archive

- The FWP should emphasize and support the need to provide access to reports and other regionally significant scientific literature in paper and electronic form.