

Collaborative Information Management to Support Ongoing Assessments of VSP, Hatchery, and Tributary Habitat Effectiveness for Columbia River Basin Anadromous Salmon

Through the Columbia River Basin Anadromous Salmonid Monitoring Strategy (ASMS), the Federal Columbia River Power System (FCRPS) Action Agencies and Fishery Co-Managers have agreed to the necessary monitoring to provide data to answer key management questions related to VSP Parameters and habitat and hatchery effectiveness assessments. Performing these assessments and reporting answers to these management questions on an ongoing basis is needed to assure effective evaluation of the Federal Power System Biological Opinion (BiOp), progress toward the recovery of salmonids listed under the Endangered Species Act (ESA), and the anadromous salmonid elements of the Columbia River Fish and Wildlife Program. This document is intended to facilitate several workshops. These workshops will gather the co-managers and other key agencies within the sub-regions of the ASMS and develop assessment and data sharing strategies for meeting reporting requirements, identifying gaps in data management and sharing capacities currently limiting the automation of data reporting, and establishing strategies to close these gaps and address any limiting capacities.

To support these data sharing and assessment needs, it is critical to have a comprehensive data management strategy that provides a regional data structure design capable of moving information from data collection, sharing, evaluation to reporting for informed decision making. A successful data management strategy will require: (i) adopting compatible information system standards and protocols to allow “connecting the dots” across disparate systems from the local level all the way to the regional or national level; and, (ii) ensuring that data from different sources can be translated to mean the same thing, thereby allowing independent databases to be combined into a compatible format for regional use.

To accomplish this, the ASMS sub-regional groups with key data management staff will be gathered and asked to identify the following:

1. Protocols, methods, and assumptions that will be used to assess VSP Parameters and habitat and hatchery effectiveness on an ongoing basis;
2. Data that will be used in these assessments and what project or monitoring effort will generate the data;
3. Schedule that indicates the anticipated timing for completing status assessments on a regular basis;

4. Entity or entities responsible for the various assessments; and,
5. Gaps that are limiting the development of a data exchange network that will allow for automation of the data flows necessary to support ongoing assessments and reporting and strategies to fill these gaps.
6. Proposals and/or adjustments to current Northwest Power and Conservation Council's (NPCC) Fish and Wildlife Program data management projects and business practices to meet the above data needs.

This endeavor builds on progress made by the region on various components of data management. For example, significant progress has recently been made through completion of recommendations, guidance, tools, and strategies that have resulted from the activities of regional coordination groups, sub-regional and watershed-specific demonstration projects, as well as tribal, state, and other organizations. Furthermore, the ASMS sub-regional groups with key data management staff are making their data management recommendations based on improving the current portfolio of NPCC Fish and Wildlife Program data management projects and business practices, thus using these to build towards the needed data management and sharing tool. The relevant data management projects currently include:

1990-080-00	Columbia Basin Pit-Tag Information	Pacific States Marine Fisheries Commission (PSMFC)
1994-033-00	Fish Passage Center	Pacific States Marine Fisheries Commission (PSMFC)
1996-019-00	Data Access in Real Time (DART)	University of Washington
1988-108-04	StreamNet {Coordinated Information System (CIS)/ Northwest Environmental Database (NED)	Pacific States Marine Fisheries Commission (PSMFC)

Additionally, this process will both consider how other ongoing efforts, as listed below, can contribute data to meet the data needs for VSP, habitat effectiveness and hatchery effectiveness, as well as how the outcome of this process may assist in facilitating data needed for these same efforts. For example:

- ESA Biological Opinions and Recovery Planning reporting
- NOAA ESA Five-Year Reviews , Biennial ESA Reports to Congress, and annual Pacific Coastal Salmon Recovery Fund Reports to Congress
 - Current FCRPS BiOp RPAs, RM&E Implementation Recommendations, and Assessment reporting
 - Columbia River Collaboration - Anadromous Salmonid Monitoring Strategy
 - Individual Recovery Board Monitoring Strategies

- Regional Technical Team assessments and reporting
- Columbia River Fish Accords Data Integration Project(s)
 - Yakama Nation Status and Trends Assessment Report
 - Columbia River Inter-Tribal Fish Commission data management projects
- Pacific Northwest Aquatic Monitoring Partnership's (PNAMP) Northwest Region-wide High Level Indicators
- NPCC High Level Indicators
- NPCC Monitoring, Evaluation, Research and Reporting Framework (MERR)
- Columbia Basin Fish and Wildlife Authority's (CBFWA) Report on the Status of the Fish and Wildlife Resources in the Columbia River Basin (SOTR)
- Ad-Hoc Supplementation Working Group report on basin-wide hatchery RM&E recommendations (Appendix C)
- US v Oregon hatchery monitoring requirements
- NOAA Fisheries Monitoring Guidance

Necessary elements of a data exchange network:

While the immediate need to report population assessments for the BiOp will require an ad hoc approach based on existing data management capabilities in the participating agencies, the long term goal is to develop a consistent regional approach that will allow automated flow of data. To meet this goal a data exchange network will be explored, allowing participating agencies to publish their relevant data in a standardized format as 'web services' on the Internet. This will allow those conducting assessments and assembling the various reports to directly access the needed data.

Necessary elements for an exchange network include:

- Documentation of the data in a structured metadata format
- Data adhere to a formal quality control process
- Data are available on the Internet as 'web services'

Specific elements of an exchange network must include:

- Data Exchange Templates
- Data Sharing Agreements
- Metadata
- Standardized terminology, including data definitions
- Protocol Documentation
- Data Quality Assurance and Control
- Database Design and Administration
- Required Database Infrastructure

Currently some of the needed elements have been developed, such as the StreamNet Data Sharing Guide and PNAMP Metadata Guidance, or should be easily developed by modifying existing products such as guidance for quality assurance and quality control. Other elements are

currently in-development through PNAMP, including guidance for developing data exchange formats & implementing & maintaining data exchange networks, guidance for protocol/method documentation found in the developing PNAMP Protocol Library, guidance for developing data sharing agreements, and guidance for data stewards.

Another element that needs to be addressed as we move toward completion of the required data management functions described below, is to ensure that participating agencies and organizations are ready, in terms of their infrastructure and access to needed expertise, to publish their relevant data in a standardized format as ‘web services’ on the Internet.

Data Exchange Templates:

Similar data from multiple sources must be compatible before they can be combined and analyzed. A Data Exchange Template (a.k.a. Data Exchange Format) defines a common structure for each specific type of data to be shared over the network. The agency collecting the data can either store the data in the required format or translate the data into the common format for sharing. A collaborative process with data managers and biologists from the sub-regional workshop groups will need to develop the templates after the data needs are defined through the workshops. A number of regional and national groups already have experience developing these templates and would be useful in this process. There is also guidance available from the USEPA Environmental Information Exchange Network (www.exchangenetwor.net).

Data Sharing Agreements:

A template for a common data sharing agreement should be created so that the agencies, tribes and other organizations wishing to participate understand the safeguards and responsibilities related to sharing their data. Key components of the agreement should include:

- Purpose of the agreement
- Legal and liability framework
- Intentions of data use and access by other web services, partners and data users
- Roles and responsibilities of network administrators, data providers, and data users
- Quality, timeliness and availability of data
- Function of the exchange network

Metadata

Metadata describe the content, quality, condition, and other characteristics of data, and are sometimes referred to as “data about data”. Metadata must explain the content and format of the information, the source of the information, the methodology used to collect it, its spatial aspects and its reliability following accepted metadata standards. Guidance on appropriate metadata standard should be obtained from the PNAMP *Regional Guidance on Metadata for Environmental Data*, (Rentmeester, 2010 - <http://www.pnamp.org/metadata>). Implementation of metadata standards by partners of this effort will enhance searching of existing data sets and discovery of new data sets and are required for the exchange network. Additionally, this will

help data users understand the meaning and proper use of datasets. Metadata standards can also be used to automate workflows within organizations and across multi-jurisdictional needs, such as with VSP assessments.

Standardized terminology

Standardized terminology and data definitions are important to simplify information sharing, but, achieving consistent use of language is very challenging. This workshop described in this process will result in description of specific data needs for conducting assessments for anadromous fish and habitat in the Columbia River Basin. With this knowledge, we will develop data dictionaries to define each data element and to standardize general terminology. This is an essential step in assuring that shared data are fully and accurately understood by data users.

Protocol Documentation

The protocols and methods followed during data collection must be clearly described in the metadata. Protocols and methods should be published and peer reviewed, and the sampling framework and survey design should be described as part of the methods. Other components should include the procedures used for collecting, managing and analyzing the data and describe the expectations for interpreting and using the data. The PNAMP Protocol Library is a tool that can be used to describe protocols and provides a simple means to refer to the protocols and methods used when writing metadata.

Data Quality Assurance and Control

Data exchanged in the network should have been subjected to a quality assurance program to ensure that the data are as accurate as possible and suited to their intended purpose. Quality control can best be provided by those people most familiar with the data at all stages of the data flow. Agencies should establish a Quality Assurance (QA) program and track compliance. Field personnel should follow agency QA procedures. Funders can use contract language to require QA procedures be followed. Development and adherence to a QA plan will help ensure data accuracy and reliability in support of resource management and restoration actions.

Database Design and Administration

Data should be stored in relational databases housed and administered by organizations closest to the source of collection when feasible. The database should be made available on servers that allow all or some of the data to be accessible through web services. In cases where the resources to design, construct, and manage databases do not exist, alternative solutions should be established such as, using other organizations that have the required resources to incorporate these data and provide administrative access to client partners.

Required Database Infrastructure, including personnel

To participate in a data exchange network, agencies and other entities must have sufficient data management infrastructure to manage the data in a relational database, translate the data to the

exchange format, and host the data on the Internet in XML format as web services. Infrastructure must include both the necessary hardware and sufficient data management personnel to operate it. Current infrastructure capacity likely varies among the participating agencies; where infrastructure is currently insufficient, a long term goal should be to improve the system to eliminate the inadequacy. In the interim, it may be necessary to host data through a partner agency or regional database project.

A broad set of skills and capabilities will be required of personnel to manage data and share it through the exchange network. The range of capabilities is beyond the means of any individual, so agencies will need to consider personnel requirements carefully. Among the required personnel capabilities are:

- Database design
- Database management
- Data management
- Metadata creation
- Oversight of quality assurance/quality control procedures
- Data system administration
- Programming
- Coordination with monitoring coordinators and data collectors
- Acquisition of data from data collectors if not part of an agency data system
- Coordination with data analysts
- Bridge gaps in current agency information technology capabilities and procedures

Timeline and Schedule (see attached gant chart)