



Independent Scientific Advisory Board

*for the Northwest Power and Conservation Council,
Columbia River Basin Indian Tribes,
and National Marine Fisheries Service
851 SW 6th Avenue, Suite 1100
Portland, Oregon 97204*

Memorandum (ISAB 2013-2)

July 22, 2013

To: ISAB Administrative Oversight Panel
Bill Bradbury, Chair, Northwest Power and Conservation Council
Paul Lumley, Executive Director, Columbia River Inter-Tribal Fish Commission
John Stein, Science Director, NOAA-Fisheries Northwest Fisheries Science Center

From: Robert J. Naiman, ISAB Chair

Subject: Review of Bonneville Power Administration's Framework for the Fish and Wildlife Program Data Management

Background

In response to the Northwest Power and Conservation Council's June 7, 2013 request, the ISAB reviewed the Bonneville Power Administration's document *A Framework for the Fish and Wildlife Program Data Management: Issues and Policy Direction for Development of a Data Management Strategy and Action Plan (June 04, 2013)*. As described in the framework's executive summary, the framework is a living document that was developed with the support of the Pacific Northwest Aquatic Monitoring Partnership based on the Regional Monitoring and Data Management Structure. The framework is intended to communicate steps that are needed to improve the management of Fish and Wildlife Program data and increase the integration with other data management programs in the Columbia Basin. Specifically, BPA hopes that state, Tribal, and Federal partners will use this framework to leverage common resources to improve management of Fish and Wildlife Program data. Ultimately, the framework is intended to support the Northwest Power Act, meet Endangered Species Act (ESA) requirements, and inform resource management. BPA hopes that key elements of the framework will be incorporated into the Fish and Wildlife Program through the amendment process.

The Council's request to the ISAB indicates that this framework addresses a subset of the Council's [July 2012 Data Management Category Review - Issues and Recommendations](#) #1 and #2.

Council Recommendation #1 (see pg 15-16):

- 1) All data collected by Program funded projects must be publicly available in accordance with applicable state and federal laws.
- 2) All projects collecting data should provide user-limited access to different levels of synthesized data to ensure appropriate use of data while providing easy access to more highly synthesized data to a wider array of users.
- 3) All projects collecting data through the Program should ensure the longevity and usefulness of the collected data by using data management approaches and tools that facilitate its sharing, such as by providing comprehensive documentation of metadata and employing data stewards.
- 4) All regional data management projects should publish their data electronically on a regular basis (i.e., not a static PDF or Word document), and consider using a dynamic data-sharing system for regularly requested information.
- 5) All individual project data that are required for program evaluation and reporting should be made accessible by making these data web-services accessible or by submitting these data to a sub-regional database or regional data-management project.

Council Recommendation #2 (see pg 17-18):

- 1) Data-sharing should be focused on priority Program data, and the shared data should not be limited to non-synthesized data, but should also include the synthesized information .
- 2) Priority data for program evaluation and reporting should be accessible through regional data management projects.
- 3) The Council should focus on improving data sharing among individual projects, subregional databases, and regional data-management projects to provide program priority data at the appropriate level of synthesis to assist with program evaluation.
- 4) To inform program evaluation and reporting needs, entities with program-funded projects that collect program priority data should engage in collaborative efforts aiming to address these needs.

ISAB Summary Comments

The framework document provides a vision for a data management structure that can be integrated with other Federal, state, and tribal data systems. The goal of this structure is first, to support progress reports on Biological Opinions and other programs in a timely manner and second, to manage and provide relevant data for future programs. Two current projects (the Columbia Habitat Monitoring Program (CHaMP) and the Coordinated Assessments Projects for Salmon and Steelhead) have shown the value of coherent data management with common data protocols, data exchange templates and standards, and regional data management.

The document identifies several strategies needed for successful Program data management: (a) identify the management questions and the data required to answer these questions; (b) require development and documentation of standardized protocols with standardized data entry mechanisms; (c) develop improved data entry methods using digital data collection methods where appropriate; (d) use distributed, structured regional data repositories such as CHaMP, PIT Tag Information System (PTAGIS), Regional Mark Information System (RMIS), Genetic Analysis of Pacific Salmon (GAPS); and (e) develop standardized reporting mechanisms. A key strategy is that data will be publically available. In general, the strategies proposed are coherent and well designed. Appendix B is particularly helpful in providing an overview of where data are currently stored for projects in the Columbia Basin. While Figure 3 provides a conceptual framework for regional data management, a master index (similar to Table B.2) will be needed to help guide users to the relevant data and show them how the different sources for the same (apparent) data differ.

Past ISRP Data Management Reviews

The ISRP reviewed databases funded through the Columbia Basin Fish and Wildlife Program in 2000 ([ISRP 2000-3](#)) and also reviewed data management projects in 2012 ([ISRP 2012-6](#)). Both reviews identified several areas of concern, some of which the current proposal addresses. For example, past reviews raised concerns about data timeliness, lack of data standardization, quality assurance/quality control (QA/QC), and coordination among the various agencies. However, some issues appear to be unresolved. For example, many ISRP reviews ask proponents to discuss emerging limiting factors. There does not seem to be a mechanism to look across the many projects to capture common emerging limiting factors and develop suitable data capture and management systems.

The two ISRP reports were also concerned about coordination among the tribes, states, and other agencies in data collection and management. This issue continues today with some agencies and programs (for example, the U.S. Army Corps of Engineers' Anadromous Fish

Evaluation Program and the Lower Snake River Compensation Program) that fall outside the BPA-funded projects and seem to manage data in isolation.

BPA's current framework document needs to include a review the previous ISRP/ISAB recommendations on data management to identify areas of concern that have not been addressed.

Specific Comments on the Framework

(1) Identification of Program-priority data and metadata

The document lacks (and needs) clear identification/definitions for the Fish and Wildlife Program priority data and metadata to be managed under this framework. Appendices A, B, and C list types of priority data and species priorities. However, these priorities need to be specifically identified and discussed in the main text of the action plan.

(2) Costs of Data Management

While the costs of physical acquisition and storage of data are declining, the costs for data management are not trivial. The framework recommends a much more formal approach to data stewardship by participating partners but makes no recommendation on how this should be funded. A sufficient portion of a project's budget should be devoted to data management. For example in the National Science Foundation's Long Term Ecological Research Network program, the percentage at one time was 20-30%. It may be beneficial for future project proposals for BPA to create guidelines on expected costs for data management functions. There is also a concern about developing capacity within the agencies for effective data management. Clarification is needed on which aspects of data management and infrastructure to support this capacity are BPA fundable. A formal funding mechanism that addresses these issues should be described and developed through the framework effort.

(3) Data availability

The Columbia Basin Fish and Wildlife Program has a policy that

“... Data and reports developed with Bonneville funds should be considered in the public domain. Data and metadata must be compiled, analyzed, and reported annually and within six months of project completion.”

The current document does not have any discussion on how well current and past projects are in compliance with this policy. Non-compliance could be related to a number of issues including lack of funding, uncertainty on where to store compiled data, and so forth. The proposal needs to review projects that are not in compliance to identify areas where more guidance in the

document is needed. The document also needs to discuss strategies to help improve compliance, for example, providing bridge funding to develop trained personnel in agencies, incentives for timely deliverables, and so forth. Who will be the final arbitrator on assessing compliance?

BPA has furthermore adopted the Organization for Economic Co-operation and Development (OECD) guidelines for access to research data from public funding (see page 14 of this report) and in Appendix D has a draft of a Data Sharing and Use Agreement. Data sharing and Intellectual Property (IP) issues are complex with competing agendas. For example, researchers may want an embargo on parts of data or data collections so that they can have priority publication. During its Geographic Review, the ISRP noted that one proponent preferred not to make data publicly available because data provided previously have been misused. Some agencies may feel that some data are very sensitive and would prefer not to release them. Because data are collected using public funds, they are publically available and so any restrictions on the use of data need to be carefully justified. Standard data disclaimers and acknowledgements of original data sources need to be developed for inclusion in publications using Fish and Wildlife Program data released to the public domain.

There are also different tiers of data ranging from the raw data, to data summaries, and final reports. Rules and ease of data access should vary across the tiers. For example, copies of final reports should be freely available on a web site, but access to individual data records may be more restricted and meta-data should be provided when the raw data are provided.

Data mining aggregates information over several different projects, but the data may not be fully compatible and so results may be at best meaningless or at worst harmful to long-term interests. The proposal needs to identify who will take responsibility for responding to data requests that span several agencies, are in different formats, and need extensive meta-data.

It is not clear from the proposal who will resolve conflicts that will arise between requestors and providers. Responding to data requests can be time consuming and again there is often no budget for responding to detailed requests after projects are completed. The proposal needs to identify who will fund these requests.

The proposal for an end user agreement (Appendix D) includes a “non-compete and sharing” clause. The proposal needs a precise definition of these terms. In any event, these are likely unenforceable if the data are in the public domain.

(4) Review of metadata

An important part of data management is the handling of the metadata. The current proposal provides a mechanism through efforts such as MonitoringMethods.org. However, oversight of

metadata needs to be better addressed – for example, are the protocols in the MonitoringMethods.org sufficiently well described as to be useful for describing the data collected rather than simply how the data were collected? Some oversight is provided by reviews conducted by the ISRP, and feedback mechanisms are available on the MonitoringMethods.org website. The framework needs to develop a more formal review process of the metadata to ensure that the metadata will be useful for at least 10 years.

(5) Historical datasets

The framework briefly notes an agreement to archive historical datasets (section 2.5.7). The presenters mentioned that Streamnet is digitizing old data, but current data have the highest priority. Historical datasets are extremely useful and efforts should be made to update and maintain consistency in these datasets, including all of the fish datasets in “Status report: Columbia River fish runs and fisheries, 1938-2001.” This now-defunct series of annual reports contained 84 data tables including some with salmon data extending back to the late 1800s. It is presently difficult to track down consistent updates to all of the datasets that were previously located in this single compilation by WDFW/ODFW. The document is no longer being produced due to budget cuts.

(6) The role of the Council

The document does not include any specific strategies or actions with respect to the Council's role in improving data sharing (refer to Council Recommendation #2). For example, Council Recommendation #2 included the establishment of a Program Evaluation and Review Committee to ensure that the data management remains focused on the Council's program needs, but the document does not indicate how this committee fits in the proposed structure. How does the Council ensure that all of the information from a funded project has been captured and is available, that the data helped inform the Program objectives, and that other programs in the Basin can make use of any relevant data for their specific objectives? The proposal needs to show how Council oversight fits into the data management framework.

(7) Data protection and security

The proposal needs to include explicit strategies and actions to protect data and to ensure data security. For example, regional data centers may have well-developed security, backup, and disaster recovery plans, but these should be reviewed periodically to ensure that they are still in force and exercised to ensure that they actually perform as expected.