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
ISAB@nwcouncil.org

Memorandum (ISAB 2006-5) **July 18, 2006**

To: Tom Karier, Chair, Northwest Power and Conservation Council

From: Nancy Huntly, ISAB Chair

Subject: Review of Council Proposal for a Columbia River Basin Data Center

As requested by the Council in May 2006, the ISAB reviewed the Council's proposal for a data management center. In conducting the review, the ISAB was briefed by Peter Paquet, Council staff, and Stewart Toshach, NOAA Fisheries, on how this proposal fits in with the overall regional data network effort and specifically the formation and efforts of the Northwest Environmental Data-Network (NED). The ISAB also considered background documents from NED and past ISRP and ISAB reports on the topic.

The Council's proposal, specifically the list of four primary responsibilities, responds directly to the critical elements to improving data availability and access identified by the ISRP in its 2000 report, *Review of Databases Funded through the Columbia River Basin Fish and Wildlife Program* (ISRP 2000-3). The needs identified in the ISRP's 2000 report remain, and it is good to see that the Council and region are seriously attempting to address them. In sum, the ISAB finds the Council proposal and NED effort supportable, but recommends that the data center approach be first initiated as a demonstration project.

Although the basic elements of a data management process are outlined in the brief proposal, there is insufficient detail to enable a thorough assessment of the proposed program. As this proposal is, in essence, a proposal to produce an RFP, more detail is probably not required. The ISAB provides comments below to help the Council develop the next iteration. Suggested edits to the proposal itself are included in Attachment 1.

Overall Purpose
The Vision and Problem Statement of the proposal are worthy and targeted on a very important problem, to improve accessibility and credibility of "all data." A process to compile and coordinate data for the Columbia Basin is an obvious need. Specifically, there is an urgent need for coordinated entities responsible for providing access to and quality control of the diverse range of environmental data accumulating in the northwest region. The proposal to create the Columbia Basin Data Center to address this need should be considered a priority, and the ISAB supports this effort. While NED and the Pacific Northwest Aquatic Monitoring Partnership (PNAMP) are making inroads, there is

an obvious need to move ahead quickly, even on a smaller scale as a demonstration project. The proposed data center could be key to making progress.

Regional Participation and Sustainable Financial Support
The development of a data center that involves so many different parties and entities needs careful thought about objectives and priorities. Before the RFP is generated, the key organizations collecting data in the basin should be in agreement that they will participate in the data center. The effort will certainly fail if even one of the major research/monitoring efforts in the basin refuses to be included in the program.

It will likely be difficult for the proposed data center to deal with inter-agency differences at the beginning. For example each agency has specific data gaps, and it would not be effective for the proposed data center to identify them – a bottom up approach may be needed since staff of the proposed data center probably could not understand the nuances of all individual agency requirements and mandates, at least at the start. Top down leadership is also needed to provide consistent support.

A show of support involves considerably more than an MOU to "share data." In this case, several "entities" have to invest resources on an ongoing basis, or it won't be sustainable. If parties operate in self-protective mode, this well-intentioned and forward-looking effort will eventually fold, for lack of support.

Someone needs to share with the Council what is needed, perhaps even identifying the target "entity" audience. How many people? How many years? How much equipment and supplies? How big an operation? How should they staff it? How is it to articulate between the primary producers and the end users? What is in it for the various players? How do we get the value added, above and beyond warehousing (portaling) data? "Build it and they will come" won't sell and is not a scientifically defensible approach. A more explicit plan for who is involved and why is needed up front in the proposal. An explicit implementation plan should include a pilot project as a first step.

Existing Databases: StreamNet
The Council proposal needs to clarify how existing data management efforts, like StreamNet, will interact with the proposed data center. Will these existing efforts be subsumed by the data center? Will existing projects continue their current data management functions and link their databases with the data center? Will there be significant overlap in responsibilities that will hamper the effectiveness of the data center? Subsuming these existing projects with a central data center was explicitly recommended against in the ISRP's 2000 report. The Council's proposal does not seem to suggest subsuming existing projects but seems oriented to addressing issues raised in the ISRP's 2000 report. The relationship of this proposal with existing efforts, however, should be resolved explicitly before releasing the RFP. Perhaps, the databases to be included in this effort should be listed in the next iteration.

Specifically, one of the ISAB's concerns with the proposed large web-based data information system is that it seems very similar and potentially redundant with the

StreamNet project. The StreamNet project is currently under review by the ISRP in the FY 2007-2009 project selection process. Although StreamNet has been responsive to past ISRP reviewers and is continually improving, reviewers have had problems in the past using StreamNet to get information due to corrupt files, the interactive map site not working, and a variety of other limitations. It is not clear from the Council's proposal that the proposed data management center will be better positioned than StreamNet to address data gaps and the need for broad access to data.

Incentives for Participation
How will the Program ensure project compliance for data entry and metadata entry? If the data in a standard format and metadata are required contractual deliverables and the need to meet this responsibility is enforced (by terminating projects that do not supply the deliverables or withholding a percentage of BPA contract funds), then the data and metadata should be supplied. If delivery of primary data and metadata is not a firm requirement, then the data center would likely be a waste of time and money. If the Council is to help solve the data existence, compatibility, access, and quality problems that have been identified, the Council and BPA must require that data gathering that they fund results in delivery of documented data for use by the parties that need it.

ISAB Recommendation: Begin with a Demonstration Project
Given the need to get something coherent and real in place and given the financial realities attendant upon that need, the ISAB recommends that the proposed data center approach be initiated with a demonstration project on a limited/pilot scale using only a specialized data set (e.g., salmon escapement data from one or two subbasins in the region). A demonstration project would prove the value of this proposed program to technical experts, policy makers, agencies, and the public in the region.

Therefore it would be best for the proponents, perhaps under the auspices of NED and PNAMP, to agree upon a pilot-scale project to test out the data center concept. As mentioned above, salmon escapement data might be a priority since these key data are not Internet accessible but are used by so many agencies in the Columbia River Basin.

This "demonstration of concept" should go two steps farther than showing that data can be warehoused electronically and with a web portal. First, the pilot project needs to demonstrate that it can coordinate with the data generators in workable fashion to bring in coherent data from multiple sources, or provide access to those data in "standard form." Second, they need to demonstrate that they can feed an "end user group" for productive analysis, so "emergent product" of value comes out of using this portal. If a pilot project can demonstrate that it is useful in this way, the enterprise will be given credibility and buy-in should be promoted.

What this entails is that the proponents team up with both a data generator group and a data user group (in advance) allowing them to carry a finite (but meaningful) problem through from data generation to data warehousing to data mining to a valuable conclusion, productively. The next stage should involve a *serious* "pilot project," providing that "proof of concept."

Some elements and issues to consider in setting up a pilot project include:

- The proponents need to sit down with a set of willing players, people who are prepared to develop that pilot, choosing a meta-problem that needs serious attention, is circumscribable (not too many players, perhaps a dozen at most), and then plan it out, start to finish, from generation of the data to the "overview report."

- All players should be in the planning and execution loop from the outset, so that it goes off with a minimum of "hitches" and that at the end, the product is valuable for planning and for the data generators (so they will keep producing data and cooperating).

- What data are needed? Who has them? Before initiating the pilot project, the proponents should make sure the cooperation they need will be forthcoming. The proponents should show how the enterprise-level effort of the data center helps (without getting bogged down in a huge effort). The data should then be warehoused or portal-connected, making sure the "end user" has access (and commitment) to the product. This idea (a group figuring out what data are needed) responds to the ISRP's 2000 report comment that direction to seeing that NEEDED data are gathered and archived is an important and missing responsibility and to the call for some top-down leadership in addressing the problem by defining data needs. This approach also belongs as an element in the Council's developing M&E Plan, which should discuss and coordinate explicitly with any data warehouse functions.

To some extent, the current database center efforts at the Fish Passage Center and DART have operated at this level, and lessons from those projects could inform development of a pilot project.

Specific Comments on the Proposed Four Primary Responsibilities
1. Internet access – a process is required to develop the standard protocols
"to ensure that data from different regions and from different sources are compatible."
This task is relatively easy for data such as water quality which can measured with instruments but might be more difficult for ecological observations where some subjectivity can be involved (e.g. habitat description). Taxonomic issues such as changes in species names also need to be dealt with.

2. Data gaps – see above comment.

3. Data integrity – the oversight function is important but might be difficult to achieve if intrusion into agency data is required. A lot of goodwill might be needed.

4. Data standards - see also comments in 1 above. Some of this work may have been done already and the proposed data center would have to avoid duplication (e.g., EPA water quality standards, PNAMP beach seining protocols).

Attachment 1: ISAB Suggested Edits on the Council's Proposal for a Columbia Basin Data Center

Suggested ISAB edits are incorporated into the Council's proposal below in blue font and square brackets, e.g., [edit].

Columbia Basin Data Center

4/24/06

The Vision

The development of a Columbia Basin Data Center will provide extensive benefits for the entire region. It will provide a much needed outlet for comprehensive, high quality data that can be used by [federal, state, tribal and non-governmental] policy makers, technical experts and by the general public. It will also work with the Council and Bonneville to ensure that data gaps are filled, standard data protocols are followed, and data quality is maintained [to facilitate evaluation of the status and trends of fish stocks and habitats in the basin.]

Problem Statement

In May 2000 the Independent Science Review Panel wrote a report on the inadequacies of the data system for Columbia Basin fish and wildlife. They noted that there were significant data gaps and significant inconsistencies in the way that data were collected and reported. They recommended a systematic approach to address a wide variety of tasks including an inventory of existing data, a survey of unmet data needs, proposals for filling data gaps, and development of standardized data collection and reporting protocols.

The Council's 2000 Fish and Wildlife Program addressed many of these same issues. The Program promoted the new vision that necessary data would be collected in a standardized fashion and made available through the Internet. First, the Program called for, "The Council to initiate a process for identifying data needs in the basin, surveying available data, and filling any data gaps." It was anticipated that the Council would adopt a set of standards that would then guide data collection. To this end the Program stated "The Council will initiate a process involving all interested parties in the region to establish guidelines appropriate for the collection and reporting of data in the Columbia River Basin." Pursuing this concept, the Program stated, "The methods and protocols used in data collection must be consistent with guidelines approved by the Council." Finally, all of this information needed some place to go, so the Program also stated, "The Council will initiate a process for establishing an Internet-based system for the efficient dissemination of data for the Columbia Basin." [p. 33]

Many of the same problems first identified by the ISRP in 2000 persist today. There is no comprehensive data inventory available for the Columbia Basin that [facilitates syntheses and evaluation or] identifies data gaps, and no projects have been approved or

even proposed to close those gaps. The Council has yet to approve guidelines for collecting or reporting data and there is still no Internet access for broad categories of Columbia Basin Data.

Through voluntary, collaborative efforts we have reaffirmed the problems identified in 2000 and verified a broad-based interest in resolving these problems. In particular the Northwest Environmental Data network has coordinated a broad variety of parties to work on these issues. The following proposal builds on those initiatives and moves us forward in resolving them.

The Proposal

The current effort to resolve these issues has benefited from the broad representation of numerous agencies and tribes. It has, however, suffered from the absence of a single entity with the responsibility and resources to move forward. This entity, with the working title of the Columbia Basin Data Center would be charged with ensuring that important data necessary to understand the status [and trends] of fish and wildlife in the Council's program are adequate and available. The Data Center would not be responsible for collecting and compiling data. That function would remain the responsibilities of project sponsors and fish and wildlife managers. No analysis would take place at this Center. The Data Center would, however, be responsible for ensuring that the full array of important data--for hatcheries, harvest, hydro passage, and habitat--are available, reliable and adequately documented. The Data Center would work with the Council and Bonneville staff [and other agencies] to identify and remedy shortcomings in the current system.

The Data Center will have the following primary responsibilities:

1. Internet Access. It will be responsible for maintaining a high level web site that would serve as a portal for existing data. This site would be user friendly for policy leaders, technical experts and the general public. It would rely on standard protocols as necessary to ensure that data from different regions and from different sources are compatible. And it would offer sophisticated web based tools for graphing, mapping, and consolidating data.

2. Data Gaps. It will be responsible for conducting inventories of existing data and determining the existence of data gaps. It would be their responsibility, in consultation with various entities in the region, to facilitate approaches that would resolve gaps. Unresolved data gaps and proposals to resolve them would be reported to the Council and BPA.

3. Data Integrity. It will provide oversight over data quality, ensuring the integrity of the data. It will do this by periodically reviewing the procedures used by different entities to assure data quality.

4. Data Standards. It will propose standard protocols for data collection, data reporting, and data quality[, in collaboration with other agencies,] to be considered for adoption by the Council. These protocols would be applied to BPA funded projects as stated in the Program

In order to achieve these results the Data Center must establish and maintain a reputation for neutrality and objectivity. At the same time it must continue to work closely with the many entities in the region that are users or purveyors of data. One approach for consideration would be to establish an advisory committee modeled on the Northwest Environmental Data Network.

It should be emphasized that the Data Center would by itself have no responsibility for actually collecting data in the field, a task often associated with [research and] monitoring activities. Instead, the Data Center would either be the recipient of these data or it would access them after it has been posted by others.

Implementation

In order to achieve this vision there are several important steps that need to be taken. Bonneville, in coordination with the Council, will need to convert the vision outlined in this paper into a more precise statement of work that support a request-for-proposals. This is the type of work that lends itself to a competitive, RFP process. In selecting a contractor for this work, Bonneville should ensure that it:

1. Can implement the tasks on schedule.
2. Can maintain a high level of integrity and neutrality.
3. Identify where data [were]are collected[, who collected the data and who is responsible for the data]but not publicly available.
4. Remains independent from the data collection process.

If Bonneville finds that the proposals are inadequate or too expensive they could consider developing the Data Center in-house. Bonneville developed a sophisticated project tracking system, Pisces, that promises to provide considerable value to the regional fish and wildlife effort. This new Data Center is more than information technology, however, and will require a somewhat broader set of skills.

Bonneville will also need to develop a budget proposal for this initiative. As it does this they are encouraged to develop a more comprehensive data management budget that includes all of the existing data management efforts in the region. The comprehensive budget should ensure that there are no duplications of functions and should identify efficiencies that can be gained by concentrating these data management functions in one single entity.

[Add a sub-section on initiating this effort with a pilot project; see ISAB comments in the memo above.]