

**Meeting Notes: Northwest Environmental Data-network Steering Committee and PNAMP
Data Management Work Group Meeting**

Time: 6/4/2006 from 9:00 to 4:00

Location: Columbia River Inter-Tribal Fisheries Commission, 729 NE Oregon St, Ste 200, Portland Oregon.

Participants: Bruce Schmidt (StreamNet), Tom Pansky (BPA), Peter Paquet (NPCC), Tom O'Neil (NWHI), Stewart Toshach (NED), Phil Roger (CRITFC), Roberto Morganti (USFS), Michelle Hollis (Port of Portland), David Tetta (EPA), Burney Hill (EPA), Greg Seiglitz (OWEB) **On phone:** Jen Bayer (USGS), Cedric Cooney (StreamNet-ODFW), Joy Paulus (WA IAC).

A. NED AGENDA ITEMS

1) Next meeting. It was agreed to cancel the planned July 5th meeting. The next meeting will be held at the Northwest Power and Conservation Council meeting room in Downtown Portland on August 2nd 2006 from 9:00 to 4:00. An agenda and other meeting information will be sent out ahead of the meeting and posted on the NED web site at <http://www.nwcouncil.org/ned>

2) Short Updates from Workgroup Leads:

Data Discovery and Sharing (Tom Pansky)

ESRI Toolbox version 3.0 is being loaded to the NED Portal site. The test environment should be ready by mid-July and the production environment by the first week of August. After testing the next step is roll out and a plan is being developed for that effort.

SubBasin Planning, (Peter Paquet and Phil Roger)

Phil has 3 more contacts to follow up on with stream information. There is some uncertainty about what versions of data sets are the most recent. There are difficulties replacing a new data set in its entirety and reattaching the fish population data to the stream network.

For the EDT Stream Reach data the entire data set has been recovered from Mobrand. The main focus is on the Fish Population data which is currently in an SQL database. Goal is to migrate it to a different format.

Phil is also writing up a critique that documents missing data and or missing files. The entire subbasin data collection effort cost \$15M so it is considered to be important to look at lessons

learned and evaluate the data collection effort. Phil will share a draft of the critique with the NED SC when it is available.

Effort is also beginning on how to look at the data from a Provincial scale perspective. A stage I analysis is complete, with some editing to be completed by the end of July.

Once all the EDT data are located there is a further issue about where they should go. Stream reach editors may be needed. Also decisions are needed on how to get the geo-referenced data to the NED portal. An important next goal is to archive all the baseline data runs, any diagnostic runs and all GIS files.

Upland Habitat, (Tom O'Neil)

Habitat Classifications:

The compilation of habitat classification systems is still in the works. Crosswalks have been completed on over 2100 habitat classification entries. For classifications that had GIS data layers depicting their coverage, Tom has overlaid these with the WHR Habitat Types data layer and tabulated percent area shared by each habitat type. This made cross-walking for these classification systems much quicker for EPA Ecoregions in Oregon, Washington, and Idaho. Approximately 1000 of the 5550 entries are unlikely to be cross-walked to the Wildlife-Habitat Relationships classification system. There are over 2000 Forest Service Ecoclasses that are based on vegetation, which are still being cross-walked to habitat types based on dominant and subdominant vegetation; 25% of these Ecoclasses have been cross-walked. Tom has also received some additional data that he has added to the list from the Idaho GAP program and a CD with a data layer for this system which will allow quicker processing of the crosswalk for this system. Tom has set up a meeting with Loren to go over importing into MS Access for the beginning of July. The project is on track for completion by July 31.

Subbasin Plans:

Data from KWA still needs to be sorted through. As it stands, KWA originally was going to give us data for 5 subbasins (they said they worked on all 5), but the data sent only covered 2. Further investigation by CRITFC revealed that KWA only has data for 2 of the 5, and they either never did work on the other 3, or they just do not have data for them. Tom has not made much progress on this project in the last month. He has records for all of the maps in each of the subbasin plans and has matched GIS data layers to these maps when possible. Several searches have turned out to be dead-ends. He has not been able to locate the Elk data layers from the RMEF, and many of the bird maps are only available as jpeg images. Twelve subbasins have been archived as much as possible. Downloaded data from online sources for approximately ten other subbasins and the data still needs to be matched.

Water Quality (David Tetta)

The main issue for the water quality group is how to make the data inventory into a dynamic web enabled inventory - potentially using a Wiki format. The Inventory is considered a good source of data and other material for posting to the NED Portal site. Some further research is needed on how many of the data sets are consistent with National Water Quality Data Standards and how these standards can be deployed by the Council through Columbia Basin efforts. An ongoing issue is how to support these tasks. The water quality group can use some staff support for these follow up tasks. Is there in-kind support?

Spatial Temporal (Stewart Toshach for Joy Paulus)

Stewart reported that the “Best Practices for Reporting Locational and Time Related Data” had been discussed with NOAA Executive who plan to deploy them as minimum standards within the Fisheries Science Center and the Northwest Regional Offices. They have also referred a copy of the Best Practices to the PNAMP Steering Committee for consideration, endorsement and subsequent adoption by PNAMP participants.

Salmonid Data Management Group (Stewart Toshach)

Three efforts are underway:

Developing consistency between the PCSRF and the PISCES data definitions for reporting project compliance. A cross walk comparison of the data definitions is complete and is waiting for executive action to work on resolution of the many differences.

Developing more detailed data definitions for project reporting for the purpose of developing effectiveness monitoring strategies. Stewart has worked with Katie Barnas and Steve Katz at the NWFSC to develop a draft “*Data Management Design for Regional Project Tracking to Support M&E*”. The document would be discussed with the PNAMP Effectiveness Monitoring group.

Completed a Statement of work: *NMFS Salmonid Status and Trend Data Collection Work Process*. NOAA staff has now completed input into the design and the SOW would now be circulated to external entities for potential collaboration on the effort.

3) Discussion on the NPCC proposal to develop a Columbia Basin Data Center.

It was agreed that the NED Steering Committee would respond as a group, with individual entities being able to provide separate comments at their discretion.

The purpose of the Council proposal was open ended to explore the potential and interest in taking this action now as a part of the 2007-2009 funding cycle.

The following comments were offered during a long discussion:

- Make some suggestions about alternative directions if these could be developed.
- Add to the history to explain in more detail the role that NED has played.
- Offer solutions.
- Make a direct response – development of a complete counter proposal was not possible in the time available.
- Refer to the White Papers developed at the Regional Data Workshop.
- Develop a short response.
- Offer that the NED SC has the ability to and has begun the process to draft solutions.
- Keep the focus regional and broad.
- Actions beyond the Fish and Wildlife program are critical for a regional data management system.
- Try to resolve questions of analysis, synthesis or interpretation of data. How much of this should happen at a Data Center?
- Keep comments at a high level.
- Include a forward planning component – it is missing from the region at present.
- Discuss “roll up” issues.
- Discriminate between management of raw data and management of derived data.
- Include opportunity to relate environmental data to socio-economic and human health data.
- Support data integrity.
- Need to address “real” problems, what is missing. Has the coordination function been resolved, what entities have any authority to take action and make decisions about regional data.
- A fish and wildlife data center is not sufficient by itself for example how to relate fish and wildlife data to human health and economics.
- Respond to what needs to be done. Deal with the how and who later.

- Need an entity to push these efforts forward.
- A data center would need to be independent and the BPA is not seen by all as independent.
- Data mining – don't do it at all.

Lunch break

- There are alternatives to RFP's including public private partnerships. There are also obligatory arrangements (eg standards) and or incentives. How to you make data sharing attractive.
- Relate the effort to a current policy initiative for example Monitoring and Evaluation. Demonstrate that data management can help with a current policy challenge.
- Most important message: Change to a more organized and consistent approach will take time and it needs to be "grown."
- Council and State agencies all recognize the need to support data management solutions but have not worked on how to get together to fund it.
- Participation will depend on the dollar cost – need to break it down into expedited steps and not take on the whole enchilada. Where is the best starting point?
- Incentives need to be developed.
- Further develop data reporting on BPA projects. Information is critical to the whole M&E effort and many groups have an interest in using core data for example CBFWA's proposed *Status of the Resource report*.

Agreed that Stewart would draft a response and circulate for comments before finalizing and forwarding to the Council before the deadline on June 23rd (See attached copy).

4) Columbia Basin Trust - Proposed Portal

Tom Pansky asked Peter Paquet to explain what he knew about the proposed Columbia Basin Trust Portal. Peter said the discussion was at an early stage with a desire to use the NED Portal to help meet the data sharing needs of the Columbia Basin Trust.

B. PNAMP DATA MANAGEMENT AGENDA ITEMS

1) PNAMP Retreat. Data Management Issues for the PNAMP 2007 Retreat on June 14-15.

More clearly define roles to avoid overlap and confusion: the PNAMP role is to solve biological monitoring issues while the NED role is to solve data management issues.

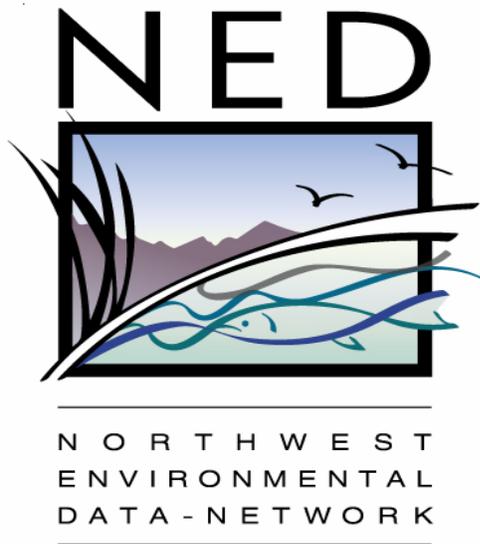
PNAMP still needs to identify what its detailed data management needs are. Typically this would require a data analyst to support identification of actual data management tasks, needed skills and resources. Previous efforts to develop this role in PNAMP were met with push back on the estimated cost of \$130-140K. Alternatively the PNAMP Steering Committee could provide in-kind resources for this task however this had been slow developing. PNAMP has adopted the SAIC report as a basic framework for needed directions in data management so it does not need to revisit high level data management needs

Another issue in PNAMP is that most of the workgroups have some data management effort underway – however each group is doing it differently.

C. AFTERNOON SESSION 2:30 through 4:00

Technology and Data Discovery Work Group:

- 1) Discussion of Data Partnership Agreement drafts. At the May NED meeting copies of a *Data Partnership Agreement Draft* and a *Portal Channels and Stewards* document were provided for discussion and the development of comments. There was follow up discussion on these documents and agreement to finalize them for use in deploying the NED Portal. Stewart agreed to make changes as discussed and post to the NED web site.
- 2) Status of Portal and Next Steps. See above for an update. The work group will develop a demonstration plan and a deployment plan as a part of the next steps
- 3) Portal Directory. An important part of the Portal is the directory structure that underpins the organization of Portal documents and provides a foundation for organized queries. There are many different directory structures. One approach would be to mirror Geospatial One-Stop. The advantage is consistency. The disadvantage is that topics of regional relevance can get 'lost' or remain unrecognized in a nationally-driven structure. Another approach is to modify the Geospatial One-Stop Portal to reflect our regional needs. Tom Pansky will discuss the options with Lenora Oftedahl who has already given this some consideration.



NED Steering Committee Response to the Council “Columbia Basin Data Center 4/24/06” Proposal

The NED Steering Committee (NED SC) has reviewed the Columbia Basin Data Center (CBDC) proposal and has the following comments:

1) Add Background Material to the “Problem Statement”

The “Problem Statement” for the Council proposal should recognize the completion of specific Council studies addressing data management needs and progress made under the existing NED Memorandum of Understanding (MOU) as follows:

In May 2000 the Independent Science Review Panel Review of Databases Funded through the Columbia River Basin Fish and Wildlife Program identified inadequacies of the data system for Columbia Basin fish and wildlife data. 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.

In addition the ISRP noted, “there was no need to centralize the entire data storage and access system. Internet Technology allows...for a distributed access system.” The ISRP also noted that no organization presently is taking responsibility for comprehensive design of data collection in the basin and recommended the formation of a joint working group to develop recommendations.

Following this report the Northwest Power and Conservation Council (NPCC) and NOAA Fisheries (NOAAF) signed a Memorandum of Understanding (MOU) to initiate a work group and co-manage an independent data management needs assessment that was completed in 2003 by Science Applications International Corporation. The report *Recommendations for a Comprehensive and Cooperative Columbia River Information Management System* was accepted by the Council.

To move forward with the recommendations a December 2004 MOU for the Northwest Environmental Data Network was drafted and has been signed by 13 regional Partners. With the new MOU in place the Partners developed a phased work plan that was approved by the Council and others.

The NED Steering Committee has also been proactive in convening annual regional workshops to develop and discuss needed actions. In 2005 a regional workshop attended, by 85 data managers and scientists produced *Final White Papers and Recommendations from Beyond Ad-Hoc: Organizing, Administering, and Funding a Northwest Environmental Data Network*.

The 2005 Workshop White Papers, the SAIC Needs Assessment Report, the NED MOU (all available at <http://www.nwcouncil.org/ned>) and the ISRB report are all important background materials for this issue.

Under the MOU and through the voluntary, collaborative efforts, NED has verified a broad-based interest in working together to develop solutions to what are complex and challenging problems.

2.0 Scope

2.1 Scope – Geographic

Scope is an important part of the NED SC response. It is not entirely clear whether the CBDC is intended to improve data management for the Council's Fish and Wildlife Program (F&WP) projects, or to improve data management for the Columbia Basin as a whole.

If it is the later, it is with one exception, consistent with the approach adopted by NED, as outlined in the NED work plans and strategy. The exception is in recognizing that adopting a strict basin approach cannot meet state or federal needs to move towards consistent data management on a state-wide, province-wide, or federal region-wide basis. So the scope of the NED effort is necessarily broader and relates more to jurisdictional than watershed boundaries.

The NED SC reached consensus that a broader regional approach, along the lines proposed in the NED MOU is the preferred choice. It is consistent with the findings of both the IASB and the SAIC reports. Also, the NED SC recognizes that effective regional information system work requires collaborative rather than individual program-by-program efforts and also notes that much of the needed data is not collected directly by the F&WP necessitating a broader regional solution. For at least one of the major concerns of the Council's program, the management and recovery of Salmonids, some of the needed data is beyond the strict boundaries of the Columbia Basin.

2.2 Scope - Topical

With respect to topical scope the SC is not clear whether the CBDC is intended to provide data management content beyond the Council's F&WP. For example, all fish harvest data and most of the data on the status of fish populations are collected under other programs. Also, data on human health in relation to water quality, and socio-economic data is limited within the F&WP but is likely to be necessary for the development of regional solutions. Adding reporting requirements to BPA contracts will be beneficial but will provide only partial improvements to information sharing.

3 The Need for a Coordinated effort.

The CBDC proposal suggests that current efforts *“have suffered from the absence of a single entity with the responsibility and resources to move forward”*. A strong majority of the NED SC agrees with this observation but views the challenge differently - to develop a single regional effort across multiple data programs and a broader geography than the Columbia Basin.

4 Existing Initiatives

There are many initiatives already underway, which, if supported and well coordinated, could meet all the primary responsibilities identified in the CBDC proposal (in italics below).

Dealing with these in turn:

4.1 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”

This is primarily an IT issue which was addressed in detail by SAIC and can best be deployed by small group with appropriate expertise.

NED has already researched, developed and is testing:

- A “*high level web site that would serve as a portal*” for existing data to allow users to discover and locate the data on the web (note that the NED portal is not the same as the NED web site). It will require support from data stewards to support ongoing operation.
- The site is “*user friendly*” and uses a developed technology being used nationally by Geodata.gov and the National Spatial Data Infrastructure.
- The site relies on standard protocols and can ensure that “*data from different regions and” from different sources are compatible*”. Note that “*compatible*” does not mean that the data can be automatically compared. Comparability relies on Metadata: information about how when and why data was collected must be available before data can be compared, regardless of standards used. Tools for data integration are also needed and NED proposes to develop a distributed approach. This is consistent with the ISRP and the SAIC recommendations.

Implementing and maintaining portal functionality over time is more difficult than building an Internet portal. Participants have different abilities to publish information on the web. Much important data is maintained at a local or district level where local biologists typically do not have the training, equipment or time to undertake this task. It will take an organized effort to maintain information sharing capability as the variety of participants respond and individually adapt to new technology.

4.2. “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.”

Scientists, resource managers, decision makers and other experts on data content (not data developers or data managers) are the people who must be involved in making decisions about what data are needed, to compare needed data with existing data and therefore to identify gaps. There are a number of existing efforts working to identify needed data content in the Basin and the Region: PNAMP, CSMEP, Columbia Pilot Monitoring efforts, State IMW’s, Tribal programs, the Puget Sound Partnership and cross-boundary efforts are examples.

- Once these groups have completed their own data inventories there is a further task of integrating data needs across these inventories and coordinating data collection across programs. It is important to understand that this is a two step task.
- NED is recommending the management of data content semantic information across multiple entities through the use of tools such as a regional data dictionary. This is a tool that can identify all the names, definitions and units of measure for data – to support systematic analysis and make progress towards standardization.
- The NED is testing the on-line portal to maintain a virtual inventory of data - as it is posted to the Internet. Once data providers publish metadata to the web, the inventory will be able to be queried to determine what data is collected, where and how, and what the data quality is.

4.3 “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.”*

Assuring data integrity is the responsibility of both content specialists (eg biologists) and Information Technology experts. Development and education about best data collection practices can improve the quality of the data collected. Development of software applications to screen and test data before passing it to higher levels can also improve data quality.

- Oversight of data quality is insufficient by itself. At a minimum - a set of regional Best Practices for Data Quality are necessary. Once developed Best Practices would need to be tested, deployed and maintained and the agencies doing the sampling, reporting and data analysis would need to adopt them. Education and training are key needed elements for data quality improvements.
- The NED approach and work plan has work elements for improving data quality- that can and should be addressed at all parts of the data collection cycle. Data

collectors, reporters, managers and analysts must all contribute to improved data quality.

4.4 “Data Standards. *It will propose standard protocols for data collection, data reporting, and data quality to be considered for adoption by the Council. These protocols would be applied to BPA funded projects as stated in the Program”*

Data standards for collection of data are best developed by data content specialists and there are (as mentioned above) existing groups in the region that are working on parts of these tasks. Data reporting standards and data quality standards can be developed for each of data collection or reporting method or protocol. Note that data quality solutions are not generic, but specific to each collection method. NED recognizes the need to develop and deploy standard protocols for multiple data sets and is working with other data content and data management groups to promote their use and development.

5.0 Statement of Work

The NED SC is willing to contribute to the development of a statement of work to define needed tasks for improved data management, regardless of the scope of the effort.

6.0 Some Gaps in the Council Proposal:

The NED SC believes that there are several important topics, relevant to successful regional data management, missing from the Council proposal. They are as follows:

- A need for formal agreements when adopting and using standard protocols and reporting methods.
- A need to improve data handling efficiency by automating routine tasks and shortening data pathways.
- An ability to consolidate data. Here, consolidating data means assembling disparate data into common views without manual data manipulation. A NED approach would lead to the use of a distributed approach, based on existing technologies and the adoption of agreed eXtensible Markup Language standards, to allow queries to multiple databases as an alternative to the manual collection of all data into data warehouses.
- An agreed upon structure for designing, planning and evaluating regional data programs. Currently there is no plan within the region or within the F&WP for developing and maintaining data management resources. NED can provide a collaborative forum for this function.
- The problem appears to be more complex than described in the (understandably brief) proposal:

- Many of the detailed issues can only be addressed by coordination amongst content groups (eg biologists) and data technology specialists. A long-term solution requires ongoing collaboration between these two types of expertise.
- Change at the data collection level is difficult because there is usually limited IT expertise and resources available. An expedient way to affect changes in data collection is to provide incentives, (equipment, expertise, and applications) which makes their jobs easier, as well as standards that allow data integration.
- The problem is large and complex, involving multiple programs, agencies, and different levels of information needs. Rather than a single “engineered” solution, we may make fastest progress by addressing one or a few specific needs to demonstrate success and build credibility. We also need to take the time to “grow” a regional data network.
- Management needs determine information needs and gaps. In the past, management needs have not always been clear. There is now, however, some convergence on a subset of information needs which are supported by a number of programs. These include fish population VSP parameters, and habitat conditions characterized by NOAA’s PFC report and EDT input data.
- Along with the need for access to a broad range of data is the need for tools to interpret that data. For instance, electric power managers use a variety of models to estimate future demand and balance that against expected generation capacity. Likewise, fish and wildlife interests need tools to aggregate data across a variety of spatial scales, create derived metrics of change, estimate the impacts of habitat change on fish and wildlife populations and etc. Analysis tools need to be developed collaboratively to assure a broad range of acceptance and trust in their results. Examples of needed tools are those being proposed by the Northwest Habitat Institute.

7.0 F&W Program Data Management Alternatives that Complement Regional Data Management Strategies

There are some initiatives that could be taken immediately by the Council F&WP that would complement the broader regional approach.

7.1 Adopt Best Practices for “Reporting Locational and Time Related Data” which has been completed by NED and is available for adoption at any time.

7.2 Advise all BPA contractors and internal programs that they will, in the future, be required to comply with Council approved data reporting and data quality standards, beginning with interim standards for 2007-2009 funded programs. Publishing data to the NED portal is a recommended step in making the information discoverable and shareable.

7.3 Complete a more detailed and explicit review of current proposals for funding of data management initiatives from 2007-2009 for the Councils F&WP. A deliberate is warranted. The first step is to decide what data collection programs will be supported. The second step is to decide what mix of data management initiatives will best serve F&WP and regional program obligations.

7.4 Set a deadline (suggest October 2006) for completion of a first set of regional interim data collection and reporting standards from the multiple data content efforts that the BPA is funding. When the interim reporting standards are complete, test them on a subset of key FWP projects in 2007 and apply them more broadly in 2008 and 2009.

8.0 Conclusion

While the NED SC understands the urgent need and the importance of making changes to improve data management for the Council's F&WP, it considers that doing so within a regional perspective and working through the programs that have already been developed, and in particular the NED effort to coordinate regional data initiatives, will provide the strongest foundation for the needed changes. The ISRB, the SAIC and the 2005 NED White Papers all identify the need for a strong regional solution and provide details of needed changes.

The NED Steering Committee and the Council have previously endorsed the findings of the ISRP and the SAIC studies that recommended a single regional entity to provide a forum for the development and completion of a distributed data design, data standards and other data resources. The Steering Committee is willing to provide this role in the region, however the current MOU anticipates funding and support to complete this task. The NED SC will bring a more detailed proposal to the Council if it is needed.

Alternatively, if there is a desire to proceed more rapidly at the F&WP level with initiatives that complement existing ongoing regional strategies - four actions have been outlined in Section 7.0 above. In the event that this is the preferred option, the CBDC should be developed collaboratively with regional partners and NED can provide an advisory committee function.