

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

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

Location: Conference Room: Northwest Power and Conservation Council
851 S.W. Sixth Avenue, Suite 1100
Portland, Oregon 97204
503-222-5161 or 800-452-5161

Participants:

Bruce Schmidt (StreamNet), Tom Pansky (BPA), Peter Paquet (NPCC), Tom O'Neil (NWHI), Phil Roger (CRITFC), Jen Bayer (USGS), Jim Gieselman (BPA), Van Hare (StreamNet), Ben Zelinsky (BPA), David Tetta (EPA), Lenora Oftedahl (CRITFC), Joy Paulus (WA IAC), Jen Bayer (USGS) and Stewart Toshach (NED) Russell Scranton (NOAAF) and Tom Iverson (CBFWA). On Phone: Michael Newsom (USBR), Nancy Tubbs (USGS), John Piccininni (BPA), Jim Geiselman (BPA).

Afternoon Session: Jimmy Kagan (Oregon Natural Heritage Information Center), Peter Paquet (NPCC), Van Hare (StreamNet), Lenora Oftedahl (CRITFC), Joy Paulus (WA IAC), David Tetta (EPA), Bruce Schmidt (StreamNet), Phil Roger (CRITFC) and Stewart Toshach (NED). On phone: Nancy Tubbs (USGS), Eric Lowrance (BPA) and Dana Collins (BPA).

A. NED AGENDA ITEMS

1) Introductions (see above)

2) Short updates

- Fed Caucus Data Needs. See - *Northwest Environmental Data Network- Summary of Work*. Stewart described the background to the request from the Federal Caucus through Jim Geiselman for a summary of NED activities. A draft of the Summary was circulated to NED SC members for comments before it was sent to the Caucus.
- Columbia Basin Data Center – See Council Discussion below
- International Columbia River Basin Center of Knowledge. Peter reported that the Columbia Basin Trust is working with the Council staff on what it needs and is scoping opportunities. Currently putting together an example of what the major interests are which should be available in two weeks. Stewart suggested that a time be identified when the NED Portal could be demonstrated by Tom Pansky, or others. A meeting is scheduled for Castlegar, BC in October, 2006.

- OFWIM Conference – October 16-19, 2006. Funding arrangements need to be secured before Stewart can commit to attending. It is likely that Bruce Schmidt will attend. Bruce has provided a short abstract to OFWIM. Bruce, Stewart, Jen and Tom Pansky will work on a joint presentation. Amy Langston and Neil Ward plan to attend to give a presentation on the planned CBFWA “State of the Resource” plan.
- CBFWA – StreamNet and Northwest Habitat Institute Review. Tom Iverson summarized the goals and objectives for the workshop and asked for planning input. It was agreed that NED SC members who were available would meet on 9/8/06 for a planning meeting with other meetings to follow as needed. Bruce said StreamNet was enthusiastic about the meeting, since it would be the first time that all of the CBFWA agencies has focused on talking about the data and a consensus was needed on what StreamNet can and should do. Ben Zelinsky thought that the management questions are needed and essential to underpin the data management efforts. Tom O’Neil is particularly interested in the priorities for meeting obligations under Wildlife Conservation Plans. Jim Gieselman was interested in whether OWEB and the State of Washington Salmon Recovery Board would be at the meeting – and in opportunities to work with State of the Salmon on their abundance tracking.

2) Demonstration of a MS 2003 Access Application that crosswalks a compilation of current Northwest habitat classifications to support consistent data definitions for aquatic, riparian and upland species - Tom O’Neil completed a demonstration of the Access Database application and invited comments. Comments made at the meeting included a request for the potential use of the application to be explained. Please Contact Tom habitat@nwhi.org for a copy. Jen Bayer will also distribute an invitation to comment to PNAMP.

3) Presentation to Council – Next Steps – Peter Paquet. Peter and Phil reported on a short meeting that they had with Tom Karier to give Tom some ideas about what sorts of pilot efforts NED was planning to present to the Council, and what sort of lessons had been learnt from the Sub-basin data collection effort. Tom had asked for pilots to be further developed in time for an initial presentation to the Council in November – what needs to be done, how will it be done and who will have responsibility for doing it. Peter said that in the Council’s proposal for a Columbia Basin Data Center that it was not envisioned as a physical data Center with data consolidation – but rather allocation of responsibility for overall data coordination and planning. Stewart pointed out that before making recommendations to the Council it would be important to revisit the specific recommendations that were made following the 2005 NED Data Workshop – which represented the views of many.

It was agreed to develop an ad-hoc work group to develop a proposal. Tom Pansky suggested that there might be an increased opportunity for buy-in if key agency staff were also briefed on the proposal before it was presented to the Council.

The following meeting dates, places and times were agreed on to complete this work:

9/19 NPCC, 8:00 through 10:00

10/4 CRITFC, time to be determined - during the Regular NED meeting.

10/10 NPCC, 9:00 to noon

10/19 If needed

11/1 NPCC, time to be determined - during the Regular NED meeting.

4) The use of a List Serve to support NED conversations - David Tetta led a discussion on the possible use of a list serve to provide a way to encourage discussion on NED topics between meetings and offered to work to establish a list serve and be the first moderator. The Interagency Web Content Managers Group was identified as a model list serve for consideration. Joy thought that it was tough to manage a list serve with lots of spam to filter out and delete. Discussion continued on whether the NED Portal could have a Wiki feature to distribute data to a group, to allow editing of groups documents showing redline and strikeout. Tom Pansky thought that it could do this. Jen explained her experience trying to use a "Quick Place" collaboration site for PNAMP: while it was a little 'clunky' it worked well for the coordinator – allowing group discussions to be organized and tracked and documents to be posted. However most of the PNAMP members did not use it at all so this meant that Jen had to maintain two systems of communication – the Quick Place system and regular e-mail. This involved too much work so 'Quick Place' was dropped. Bruce said that StreamNet used a collaboration tool for some but not all on-line discussions with varying results. Dave agreed, as way of trying out web capability, to look into setting up a Wiki site to review documents.

5) Draft documents for approval: *Check List for Organizing Field Collection and Management of Data*; Stewart Toshach, Joy Paulus. Stewart and Joy summarized the check list and why it was first developed. Jen asked if it could be tied to the Oakley et.al. long-term monitoring guidelines that PNAMP has adopted as a guide. Joy thought we could look at it but that it could be too narrow – since the audience for the check list was for any field data collection. Phil offered that it was a good list and valuable but wanted to know if a detailed 'how to guide' was planned that could if used establish consistency across the basin. Stewart and Joy agreed that it could be useful but that it was not a part of this particular task. If a 'how to guide' was needed the task would need to be scoped and resources identified to complete it. For now it will need be on the NED 'wish list'.

Best Practices for Data Dictionary Definitions and Usage. Stewart outlined the need for this document – that data dictionary nomenclature is at the foundation of consistent information system development, which is a core goal for NED. The document recommends that NED adopt the Best Practices for its internal efforts and encourage others to do the same. It was agreed that any comments from the NED SC should be returned to Stewart by 9/28

6) Columbia River Hatchery Review – Peter Paquet. A review is underway – similar to that recently completed for Puget Sound by the Hatchery Scientific Review Group. The study is expected to make recommendations about the collection and analysis of hatchery data and is being completed by Moberg, Jones and Stokes. Discussion in NED concerned whether there would be integration with other hatchery studies including the APRE (Anadromous Production Review and Evaluation) effort which was intended to create a complete Hatchery database product. The NED interest is to determine if there is consistency between hatchery data collection and management procedures and other data sets that would facilitate data integration, sharing, analysis and publish this data to the NED portal. In addition an inventory of hatchery facilities and operations would be beneficial to the region for data sharing.

B. PNAMP DATA MANAGEMENT AGENDA ITEMS

1) Inventory Task (Bruce Schmidt). Bruce reported that a report was in preparation to present to PNAMP on 9/7/06.

2) Training for Metadata. Jen provided an update on discussion with Viv Hutchison. USGS can provide training in the spring 2007 and is interested in discussing NED needs for a focus on metadata reporting, especially for the NED Portal. The next step is to discuss the needed training directly with Viv, and if it can be provided, identify a date and locating a computer lab. BPA has a training lab that could be used. Jen will follow up with Viv and put her in contact with Stewart.

3) PNAMP Data Management Data Analyst proposal (Jen Bayer). Jen Bayer outlined the proposal for data analyst resources for PNAMP and the contributions to PNAMP that the analyst could make.

4) NED Review of “Data Management Design for Regional Project Tracking to Support Monitoring and Evaluation”- Tom O’Neil, Joy Paulus & John Piccininni. Tom, Joy and John reported on the progress that they have made in review. Tom provided some detailed comments and an edited section. They all offered that the document could be improved by identifying: who the audience is, having it edited to make it more understandable from a less-technical viewpoint, correcting different use of tenses, and making sure that all acronyms are spelled out. In addition Tom offered the following: If NOAA is going ahead with development of the database described in the report why was it not just an internal document rather than being reviewed by a wider audience? How does the system relate to the locator proposal that the ISRP reviewed by NHI...ELVIS – Ecosystem Locator and visualization information system – there is a potential for collaboration? A clearer understanding of what is being proposed is needed - for instance, the section that addresses Design Specifications does not give specifics - only general statements. The figures need detailed attention for example, Figure 1b. is cited but not clearly marked...in the text it states... “In Figure 1b this set of criteria has been applied to the data in the John Day basin from PNSHPTD”. It is not clear what the set of criteria is. Figure 3.develops new concepts but does not state what they are and the discussion goes into Ecological Pathologies...will management activities effect either habitat correlates or species? It is unclear what an ecological pathology is. It would be best to avoid generalities - are we really talking about effecting climate change in these ecological pathologies? Is climate change going to be monitored? The final tables need to be modified to add a wildlife perspective – currently they are focused only on fish. John Piccininni was interested in how the project would relate to other data projects in the region – how does it fit in and would the data be publicly accessible?”

C. AFTERNOON WORK SESSION for NED Portal Technology for Data Discovery and Sharing Work Group (Tom Pansky) 1:00 through 4:00

Tom Pansky announced that the NED Portal was now available on the web as a Beta version. The URL for the Portal is as follows: <http://nppc.bpa.gov/Portal> (Users need to remember to turn off their ‘pop-up blockers’). The Data Delivery Extension (DDE) has been added to the Portal enabling “clip, zip and ship” functionality.

Tom identified main remaining steps to populating the Portal:

- Adding the Metadata records – Prepare the Metadata in FGDC or ISO standard format which the Portal will accept. You can either upload metadata in xml or use the Portal template to enter your data directly. The initial goal is to have all NED members publish their metadata to the Portal.
- Making your data available to the Portal as a web map service or a web feature service.
- It will be also necessary to provide links to the Northwest Environmental Data Network Portal Channels and Data Steward Roles and Responsibilities document and the Northwest Environmental Data Network (NED) Portal Data Sharing Agreement
- Resolve business issues and operational requirements (see below)

The Portal can also be set up to automatically harvest data from contributors based on defined geographic coordinates. . Different users can be granted different access rights.

1) Developing a plan for 'populating' the site. There was considerable discussion on this topic and two schools of thought – one idea is to harvest all existing data from an existing source, such as geodata.gov (the low hanging fruit)...then add regional data later. The other idea is to limit data, at least initially, to data sets that are from known regional sources. The goal here would be to work towards establishing the NED Portal as a source of quality data, rather than all versions of all data. There was a consensus that there is no right way to do this – that the first step should be to harvest from all NED members and then review progress.

There is a 'duplicate record' problem to solve –if we automatically harvest data there is not an easy way to cull the data out later if it is not suitable – which was raised by Van Hare and discussed in detail but without a clear solution. The experience at WAGIC, as reported by Joy, was that it took a lot of time to get about 150 data sets populating the Washington Portal on a data-set by data- set basis.

Jimmy said that the Oregon policy is to make all data sets available including county and other data that is collected using widely different standards/protocols. Jimmy also thought that there would be a benefit for harvesting data from all state clearing houses. This would provide a view of data sets that was otherwise unobtainable- and there was interest from PNW-RGIC to have this capability. Stewart thought there might be different approaches for different data sets with different groups being able to manage their particular data interests according to the philosophy that suited them – and matched the resources they have available to do the leg-work. The data Steward can give authority to harvest or publish data. It is important to develop the capability to provide feedback on data and to provide data pedigree and there is understood to be value in doing this. Nancy thought that the users could decide if they had sufficient information.

2) Engaging a NED Portal Channel Steward.

A description of the tasks has been prepared and finalized. There was discussion about whether this was a full time or part time task. There was consensus, given the size of data holdings and the pilot nature of the effort that it was reasonable to start with a full time position for a year and then

review. The next steps are to develop a request for funding. This can be done as a part of development of pilot proposals for the Council.

3) Finalizing the Portal classification system.

The challenge that Oregon faced when developing a system for the North Coast Explorer was to provide a system that worked both as a navigation tool (for those who don't understand the topical context) and a topical index (for those who already know the names of the data they are looking for). Lenora Oftedahl provided a draft Index, based on the Oregon system – and ISO Standard and other recommendations (see **ATTACHMENT** below). Again there was extensive discussion...and we looked at both the Index that Lenora provided and the geodata.gov site. A final decision on what index to use was not made, however there was agreement that we did not need a perfect system and that it could be changed later. Stewart recommended that, in the absence of an alternative we begin with the draft from Lenora. Jimmy Kagan will forward his latest version to Lenora who will work with us to make changes.

4) Identifying "data management" news topics and articles for the Portal home page – Stewart volunteered to work with Tom and Peter to draft a 'launch' page for the Portal, and to work with Eric Schrepel to provide a link from the Council's NED web site to the Portal.

Summary of actions/decisions for Portal Deployment:

- Decide on use of harvesting tools and roles
- List of targeted providers of metadata or data – start with list from NED water quality work group, access to Forest Service Data (through Federal Caucus), and ICBEMP.
- Define geography – for harvesting
- Seek funding for Portal Data Channel Steward position
- Finalize the Classification system

NEXT MEETING WILL BE AT CRITFC ON OCTOBER 4th...and agenda will follow.

ATTACHMENT: NED Portal Index and Related Recommendations From Lenora Oftedahl (CRITFC)

I highly recommend the approach used by Chesapeake Bay and create a separate interface for the different types of users. The NED portal should become a one-stop shop for all users interested in the Columbia River Basin. The NED portal would be the online version of what Dr Tom Karier recently proposed in his recent paper on the Columbia Basin data center.

- | | |
|-----------------------|---------------|
| 1. Plants and animals | 2. Fish |
| 1. Wildlife | 1. Anadromous |
| 1. amphibians | 2. Resident |
| 2. birds | 3. Non-game |
| 3. mammals | 4. Marine |
| 4. reptiles | 3. Plants |

- 1. Trees
 - 1. Deciduous
 - 2. Coniferous
- 2. Grasses & Forbs
- 3. Flowering
- 4. Mosses
- 4. Insects & Invertebrates
- 5. Fungi & Lichens
- 6. Algae & Bacteria
- 7. Threatened & Endangered Species
 - 1. Plants
 - 2. Animals
 - 3. Invertebrates
- 8. Invasive species
 - 1. Plants (weeds)
 - 2. Insects
 - 3. Fish & aquatic invertebrates
 - 4. Wildlife
- 9. Biodiversity
- 2. Habitats & Vegetation
 - 1. Vegetation
 - 1. Forests
 - 2. sagebrush & steppe
 - 3. grasslands
 - 4. historic vegetation
 - 5. potential natural vegetation
 - 2. Freshwater
 - 1. Rivers and streams
 - 2. Lakes & reservoirs
 - 3. Marine & estuarine
 - 4. Wetlands & Riparian (how does riparian differ from Rivers & streams above?)
 - 5. Wildlife habitats
 - 1. Upland
 - 2. Riparian
- 3. Water & Air
 - 1. Water quality
 - 2. Water flow & quantity
 - 3. Rivers & streams (does this duplicate as another point of finding this info, or is it different info? If different, this should have a different name)
 - 4. Lakes & reservoirs (see note above)
 - 5. Marine
 - 1. Marine (deep water? Continental shelf?)
 - 2. Intertidal
 - 3. Estuarine
- 6. Air & Air quality
- 7. Watersheds & hydrologic units
- 4. Geography & Geology
 - 1. Geography
 - 2. Geology
 - 3. Geomorphology
 - 4. Soils
 - 5. Hydrology
- 5. Land & People
 - 1. Agriculture & farming
 - 2. Forests & forestry
 - 3. Land use
 - 4. Outdoor recreation
 - 5. Sustainability & conservation
 - 6. Climate change (or under Threats & Risks?)
 - 7. Earthquakes & Volcanoes
 - 8. Fire & fire risk
 - 9. Floods & tsunamis
 - 10. Landslides & coastal erosion

These are all interesting subjects/topics, but how many are we actually going to have data for from the very beginning? From a design standpoint, rather than being inclusive from the get-go, better to have only populated subjects in the menus. That way users are not going to get blanks when they are randomly clicking through the topics. The Portal could put up new topics as they are populated with a lovely little 'new' button next to each one.

Also, many of the topics in the list seem duplicative and without further elaboration, users may be confused about where their needed data is located. This situation will be exacerbated if all the data on Rivers & Streams are not duplicated under each of the topics. While links are easy to duplicate, do we really want to have that much information listed on these pages?

As a portal, the NED site should be constructed to include the ISO definition for the topic, other references that illustrate each of these topics, and then tie into the data that has been collected and submitted to the portal for publication. In this way, the NED portal also serves as an educational tool for users.

Metadata should be a requirement for all data submissions. “Metadata is a key component of the geospatial data set. It carries critical information as to the dataset purpose, location, content, and lineage. Geospatial software and analysts increasingly rely on metadata to ingest, display, and manage data. Perhaps most significantly, metadata is the consumer information needed by a rapidly growing geospatial data market to locate available geospatial data resources and assess their fitness for a particular use. Metadata instills data accountability and limits data liability. The timely capture of metadata is fundamental to the quality of the data set as a whole. (Wayne, Linda. Institutionalize Metadata Before It Institutionalizes You. November 2005)

The lack of metadata, however, should not prevent the acceptance of a data set. Rather, the data set can be accepted provisionally and provided to select parties until the metadata is completed by the submitting party. If the submitting party is not able to complete the metadata, the task can be given to professionals to try to fill in the blanks. Unfortunately, third party metadata will never be as complete as metadata provided by the original generator.