Data Sharing Workshop to Support Coordinated Assessments

October 5th and 6th, 2010 - Portland, Oregon

Summary of Key Themes and Action Items

The goal of this workshop is to develop a more efficient process across the region to share the large amount of data requested of entities working on the ground to support regional reporting.

Staff from BPA, NOAA, NPCC, and PNAMP provided an introduction and background for their participation in the workshop. This workshop is just the first step in a much longer process and will help identify how these three agencies can help support local entities in making their data regionally available. These entities highlighted the following:

- As follow-on to the Anadromous Salmonid Monitoring Strategy Workshop ("Skamania") in 2009, we need to now identify priorities for data management so that we can align resources better and start to chip away at data sharing needs to support regional reporting, from several funding sources. This workshop should help us set our direction in doing that.
- We know that different users have needs for different levels of information and that, for example, these high level indicators may not be your current priority nor do they represent all of the data that is needed for regional reporting. They do represent a significant portion of the data we need, and sharing data to support these indicators will be a monumental step towards improved regional data sharing.
- An objective of Coordinated Assessments and this workshop is to make our data management (DM) processes work better at serving both local and regional needs.

Overview of Coordinated Assessment

Tom Iverson (CBFWA) provided an overview and background information on the Coordinated Assessment effort. His presentation is available at: http://www.pnamp.org/node/3084.

Tom Cooney (NOAA-F), Dave Ward (CBFWA), Brodie Cox (WDFW), and Jim Geiselman (BPA) described recent regional reporting efforts and their entities' experiences and needs for regional data sharing.

The remainder of these notes focuses on the three key action areas for workshop follow-up:

- The Data Exchange Template;
- Data Analysis Diagrams; and
- The Gaps, Needs and Priorities Review.

These three action areas support the workshop's primary objectives:
Identify needed actions to discover and inform partner and regional data sharing needs and priorities to:

a. *Advance recommendations for regional standards to improve data sharing and support basinwide assessments.*
b. *Inform optimal allocation of limited, regional data management resources, including Columbia Basin Fish and Wildlife Program funds, NOAA funds, and other regional data management funding sources:*

i. *For individual partner capacity*
ii. *For shared infrastructure*

Each is discussed in greater detail below.

The Data Exchange Template (DET)

Bruce Schmidt (StreamNet) provided an overview of the DET and the development process. His presentation is available at: http://www.pnamp.org/CoordAssessments.

The group discussed the development process and identified that a "successful" DET would support the following objectives:

- Provide targets for local system developers to stay aligned with regional efforts and ensure they can provide data in needed formats
- Make routine data updates more efficient because the DET format (or a data format informed by the DET) is widely available
- Agencies and Tribes no longer need to scramble to format data for ad hoc data requests because they will have already built the tools to produce data in the DET format efficiently
- Data availability (and or inventory of existing methods) improves because the DET provides a uniform structure for data such as:
    - Data contacts and sources are easily identified
    - Variations in SAR calculations become available with a common language
- Provide a common format that supports the ability to access the data easier (via a web service, direct download, or other technology)
- Provide guidance/experience for other metrics development
- Inform our approach to a standard metadata template

The group identified five existing common data "consumers" that could use the DET to gather data: NOAA 5-yr Review, FCRPS BiOp Comprehensive Review, Fish & Wildlife Program Amendments, Accord Reports/Check-ins (related to FCRPS), and Recovery Plan Updates (specifically for the Upper & Mid Columbia).

Extensive discussions were held on the content of the DET especially the definitions of terms and options for variations in SAR. These comments were collected and sent to the DET committee.

Next Steps:

1. The DET committee will move forward with the DET and update the template based on feedback received from the pilot testing and at this workshop (version 0.8).
2. The agencies and tribes will populate the DET for selected representative populations to:
    a. Evaluate and refine the DET
    b. Provide a common target/touchstone against which the gaps and needs review work can be conducted (see below)
    c. BPA has provided funding to PSMFC to fund technicians to help with this process.
    d. The DET committee will be in touch with entities to discuss how this process will move forward.

3. The updated DET will be reviewed by the DET committee who will vet the new version through their staff to ensure the metrics collected are the right ones.
    a. Anyone interested in joining the DET committee should contact Jen Bayer (jbayer@usgs.gov) and Tom Iverson (tom.iverson@cbfwa.org) via email.

Data Analysis Diagrams

The data analysis diagrams were discussed from a number of perspectives. Several participants noted the critical role they played in documenting internal processes as a crucial step in process improvement and transition planning/training. The group also discussed the role diagrams could or should play in the larger DET and gaps, needs, and priorities assessment. The group determined that attempting to do these diagrams across all populations was impractical. The following set of "good uses" for the data diagrams was identified:

- Support internal process management and evaluation.
- Documenting changes in analysis/calculations methods. For example, these diagrams could complement the narratives produced by the TRTs to document changes to methods as they are made.
- Support DET development by depicting the most common analysis approaches and ensuring that the methods they represent are accommodated by the DET.
- Complement the gaps, needs and priorities assessment by providing a means of illustrating where in the analysis process data sharing gaps impact the ability to produce and share the three priority indicators.

Next Steps:

1. The planning group will identify the representative set of populations to request data analysis diagrams.
    - This subset of populations will be developed to ensure we have good coverage of the basic methods.
2. Evaluate the creation of templates for basic methods to help people along in the development of their diagrams.
    - There was some concern that there are so many exceptions in the methods that in most cases this process that results may not be comparable.
3. Tech resources will be provided to support diagram development for both the representative populations and other populations as desired by agencies and tribes.

Gaps, Needs and Priorities Review

The group discussed the rationale and best approach for collecting information on gaps, needs and priorities in support of making informed recommendations in the early spring 2011. The following assumptions were highlighted to help focus the review:

- This effort builds off of the regional monitoring strategy to focus on data that is/will be collected.
- Use the basic principles that data should be managed locally to the greatest extent possible, and that there should be local capacity to analyze/transform data into the required reporting/sharing formats needed. The gaps, needs, and priorities will be assessed with respect to these principles.

- The focus will be on everything that needs to happen from the point of collection to the point of its internal use or external sharing. This was not addressed in the monitoring strategy work.

The group was asked to identify the base capabilities required to make this process work (Q: What do you need at your agency to do coordinated assessments in an ideal world?). The needs identified are a starting point for the development of the questionnaire that will be used during the gaps/needs assessment phase. Participants brainstormed the following list as a starting point of needed capabilities:

- In general, follow the lifecycle of data from collection to sharing
- QA/QC at all stages (especially standard data collection template that provides easy QA/QC feedback)
- Database management and data storage
- Moving/reporting data between local /field office where it is collected and whatever more centralized databases (Agency or Tribe) is used to aggregate it
- Report applications/tools and process used to normalize, and analyze data. Taking it from the primary data in the database to the metrics and indicators
- Progress reporting (internal and external)
- Managing for consistency in methods within an Agency or Tribe
- Managing data for survey designs which require large data sets
- IT and data management operations (e.g. maintenance and backup)
- IT Staffing
- Adequate Hardware

The group also identified "soft" capabilities and challenges that impact local data management. Some of these may be usefully characterized as "challenges" in the review:

- Convincing biologists/agencies to use a data system rather than legacy spreadsheets.
- The adequacy of tools provided to biologists – if they are not flexible and are hard to work with, they will not be used. It is essential that biologists be able to get data out of these systems easily in order to do their own analyses.
- There is resistance from scientists to providing (even for internal sharing) raw data because data may be used for purposes other than project design.
    - Some of this is generational.
    - A commitment for user support and training would help with this issue.
- Defining "enterprise" data, e.g., what data must be managed in a centralized system and what can safely be distributed? In most cases data is "enterprise" if it goes into the calculation of an indicator.
- We have to establish/maintain a user perspective when developing systems – in other words, the systems developed must help staff do their job better and more efficiently.

The goal of this effort would be to complete the list of individual agency gaps and needs and develop a list of projects that could begin to address them. Agencies and Tribes could also suggest shared infrastructure projects where they see cross-cutting needs. These types of needs will be re-assessed once the compiled list is available. The agency and tribes will set their own internal priorities, as well as establish where external funding is necessary.

Next Steps:

1. The planning team will develop a short questionnaire based around these capabilities and challenges.
2. The planning team will coordinate how the technical resources can be used to support collection of this information.
3. Anyone interested in joining the planning group should contact Jen Bayer (jbayer@usgs.gov) or Tom Iverson (tom.iverson@cbfwa.org) via email. Current participants: Tom Iverson, Jen Bayer, Bruce Schmidt, Scott Rumsey, Jim Geiselman, Phil Roger, Erik Neatherlin, Lance Hebdon, Brodie Cox, Cedric Cooney, Jay Hesse, Kathryn Thomas, and Steve Rentmeester.

General Discussion of Collective Opportunities to Improve Local Data Management Capacity

In a related but separate discussion, the group brainstormed generic, but collective opportunities to improve local data management capacity. No next steps were identified for this list but these themes may be useful in evaluating the individual gaps and needs discussed above.

How do we better support local data management?

- Provide a source of stable, long-term, funding – hard to develop long-term systems when funding "shifts with the wind".
- Support coordination and consistency between the groups that are collecting the same kinds of data (i.e., coordinated and consistent local data management) or are working on similar problems. There has to be a tight "problem" focus for these efforts to be successful.
- Provide opportunities to learn and share from other agencies – (e.g., tribal data management workshop).
- Look at the ISEMP model – it sounds like it has been successful at providing a template for easy input and QA/QC. Provide resources to train people on new sets of tools (e.g., show biologists how relational database work, how to visually do something inside a database, etc). MS Access is a good tool to learn how relational databases work but is not good for storing very large quantities of data. Foster environments where staff can move past spreadsheets to more coordinated data storage.
- Encourage people to follow models of existing databases.
- Provide the ability to share existing tools.