SMALL GROUP DISCUSSIONS

Break-out groups:

- Salmon Population Data Documentation
- Data Modeling Schema
- Analytical Tools
- Semantics
- Data Visualization

Salmon Population Data Documentation

Where are we now?

Problems:

- Documentation is not preserved
- No standard standards
- Lack of management support (education)
- Lack of true benefit to the collector. No immediate benefit (perceived benefit)
- Documentation at follow up stages do not exist

Where do we want to go and how to get there?

Short term goals (6 month)

- Identify incentives for data collectors (money, etc…)
- Educate the managers
- Compiling metadata field lists that have already been identified
- Compile list of metadata creation tools and identify pros and cons
- Make sure everyone has input. (otherwise we will lose somebody along the way)
- Interviewing data collectors on their basic tasks for the purpose of identifying ways to help them save time so they can develop metadata.

Long term goals (2 yr +)

- Quality codes placed on historical vs. present data
  (so that analysis can be consistent overtime)
- Map data fields beyond data collection stage
- Document historical datasets – (i.e. data rescue)

Database Architecture

Where we are today?

Current data model/database architecture methods for storing salmon data are developed from a subjective short term standpoint. Current practices of database design are based upon the immediate need of the agency and the resulting database structure is designed to solely address that issue. When additional data types and sampling designs are adopted by the agency the existing database

structure breaks and is unable to store these new types and sampling designs. Database administrators are then required to develop new data structures to store addition sampling types and metrics leading to issues of scalability and manageability.

Metadata storage is also problematic in that metadata information is currently stored in a tabular/field structure. Adding addition metadata descriptor fields requires modification to database structure. This solution is not scalable, as more descriptive information is produced regarding a data set or sampling event the required complexity of the metadata database structure increases linearly with this descriptive complexity.

Where do we want to go?

Agreed that we want to move toward a more flexible data model (Observational Model) and storing metadata as a coherent data element Vs a series of database fields.

How will we get there?

Deploying and utilizing the observational data model on the institutional level may be met with friction from the user community. This data model is more abstract and flexible than a sample design specific data model and consequently will be more difficult for users to understand. In order to address this problem, users will need to be able to produce 'outputs' from the observational model which match the data structures that they are familiar with. This way the business practices (report generation, data analysis) of the organization will not be interrupted. In addition, developing simple views of the complex/abstract data model will ensure data will be stored in a highly flexible and scalable structure in addition to not interrupting the business practices of users.

Moving metadata into a coherent single XML database field will drastically simplify metadata storage within a database structure. Staff will have to be trained in the use of various tools and technologies to parse and query native XML documents. In addition, analysis and information systems will need to be modified to interact with this data type.

Analytical Tools

NOTE: The group focused primarily on analytical tools needed to produce data base metrics (catch and escapement by stock) for each project and not on analytical tools to use in the data base analysis and visualization steps once basic information is captured in the system.

Where are we now?

- Analytical tools are most heavily needed in the project for the BC project.
- All areas have basic escapement surveys/estimates for a number of streams.
- ODFW already uses a balanced spatial design to representatively selecting survey sites and expanding survey results for a variety of spatial scales.

- Alaska already has established algorithms and analysis steps established for its in-season management.
- DFO doesn't have an established approach for expanding escapements to CU or for allocating catches to CUs.

Where do we need to go?

- ODFW and ADFG mostly need streamlined data management approaches that incorporate existing analytical processes. DFO needs to develop the best available methods to expand survey or stream escapement numbers/estimates to CU, and then to use the best available analytical method to allocate catch by CU.
- Oregon needs to develop secondary analytical methods that could better connect data types and streamline estimates of derived metrics/parameters like productivity.
- All agencies ultimately will want to validate and improve their basic estimates of escapement and analytical extrapolations of catch and escapement by populations and population aggregates. These are typically longer-term, ongoing feedback and adaptive management steps.
- Each project needs to document analytical treatment of its basic escapement and catch allocation by stock (unit) metrics so that the derivation of reported outcome measures has a transparent basis.
- All agencies desire to provide basic data summaries or outputs on their websites that monitor performance against key outcome measures. In each case users should be able to drill down, however, to see the more detailed information that are used to produce these results, documentation of analytical procedures used, and the source and documentation of the basic data collected in the field. This promotes understanding of the information and its use – promoting transparency, credibility and a strategy for discouraging inappropriate use or conclusions of data by people unfamiliar with it.
- There are opportunities to share analytical approaches across and outside project areas to help promote most robust approaches and share knowledge, expand application.

How do we get there?

- Short-term emphasis in BC is to convene regional and cross regional work groups to review and select most appropriate run reconstruction assumptions and methodologies for pink and chum, approaches for CWT and model derived estimates of exploitation rates for coho and chinook, and using best available mix of approaches for sockeye. Procedures for expansion of escapement indices to CU will also need to be developed and applied.
- Longer-term emphasis will be validation and review of data and methodologies in order to improve allocation of field resources to improve data and its expansion/application.
- Sharing information on analytical approaches, especially once data management becomes more efficient, also will provide time needed to review the level of refinement that agencies are able to create on estimates by fine population resolution.

Note: Group discussion highlighted that a number of analytical tools are available to produce standard types of population analyses. Presumably this relates more to status and trend type presentations that would be useful to summarize and visually present output of database metrics through web applications.

Semantics

**James to provide group discussion notes*

Encode that different meanings can exist for the same term/concept/field (removes ambiguity/interpretation)
Data dictionary is a form of semantics (precursor to building an ontology)

Long term goal (ideal)
-domain ontology "salmon science" domain
-interoperability facilitated ontology

Talked about making a subdomain prototype
For abundance and productivity salmon data dictionary.

This work joins well with PNAMP monitoring glossary (lacking some of the theoretical aspects)

Data Visualization

Where are we now?

- We (ODFW) have a rich information system that we want to make available to clients: (technical, scientific, policy makers and general public)
- We have spatially explicit info for time series of: adult spawners, habitat conditions, marine freshwater survival rates, juvenile abundance and distribution.
- We have a clear set of questions that we know we want answers to: Metrics, Evaluations, Thresholds.

Where do we want to go?

- We want to provide a system which allows people to view these different datasets and analyses in different ways, at different geographic scopes. Need to be able to visualize different levels: big picture vs fine scale. Some users will want raw data and others will need actual analyses, or summarized information.
- Need a clear road map of what is available in the visualization; can't get too cluttered since there is so much available.
- Things can't break regularly or not get stale or out of date. If people get lost easily they won't go back to the site. How do we also design it so that it's flexible for changes in how the data is collected in the future? (e.g. population structure)
- The Tool needs to be able to pull data from multiple sources on the fly.
- Standard outputs for in-house and additional outputs for the public.

Where do we need to go?

Short term:

- Figure out what software and platform we can use. That may determine some of the outputs.
- Basic raw data visualization and chart production.
- Need short iterative feedback cycles. Include mock-up wireframes.

Long term:

- Have the tool calculate analyses on the fly at different geographic scopes.
- Supplemental information that should get reported but that isn't necessarily a benchmark.

Next steps…

- Coordination with NOAA and PNAMP (set-up a meeting)
- Develop mockups
- Assess visualization needs, constraints, resources
- Roadmap for iterative development (dev/feedback cycles)
- Produce charts for the different metrics;
- create template so it can be done on the fly.
- Map Production? Need to come up with a template so that this information is easily updatable.
- Beta testing in many/all phases.

***Higher priority that the pilot is transferable. More important than certain far reaching technological needs or desires.
