

**Salmon Data Access Working Group (SalDAWG)
MEETING NOTES**

Monday, November 3 & Tuesday, November 4, 2008
Chemainus Room, Coast Bastion Inn
Nanaimo, British Columbia

Table of Contents	
Presentations	p. 2
Facilitated exercise	p. 14
Small groups	p. 18
Beyond this meeting	p. 23
Feedback	p. 24
Attendees list	P. 25

PRESENTATIONS

Day 1 (11/3/08)

Introductions from Cathy Kellon, Rich Lincoln (State of the Salmon), and Mike Fraidenburg (Facilitator)

Background on the 3 Agency Partnership Initiative (API) projects:

1. Alaska Department of Fish & Game (ADF&G) – Steve Moffitt

Copper River and Prince William Sound
Data Salvage and Data Access Tools

Goals and objectives

- Salvage historical data (Escapement, ASL)
- Develop additional web-based data management and visualization tools (in order to aid decision-making and assist data-sharing)

Description of area and fisheries

Copper and Bering rivers, and Prince William Sound region
11 districts and 3 gear types, (purse seine, drift gillnet, set gillnet)
Both commercial and subsistence fisheries

Current data sets

Harvest, escapement, age/sex/ size, sex ratio data, thermal and strontium otolith marks

Current data processing flow

(for aerial data) Aerial index streams → AUC, weir sites.
Captured on paper during aerial surveys
Entered into Rbase database
Exported as CSV
Summarized with FORTRAN
Reformatted for printing with K-edit macros, 33 pages per survey, 2 species, 9 districts, 28 subareas, and 215 streams for daily and cumulative counts.
*need to streamline this process, and create a visualization
(ASL) most captured with electronic measuring board
Downloaded as ASCII and error checked...very program intensive.

Existing online tools and sites

ADFG mariner system – non public (not that useful)
ADFG zephyr database – non public not useful
ADFG ASL inventory – public but not been updated 2003. lacking sufficient resources
PWS and Copper River home pages are another data source but are the format is almost never in the form that people request.

Rob Bochenek (Axiom Consulting)

Requirements analysis summary

- Initial user interviews (customers and stakeholders) to figure out user needs
 - Develop those needs into measurable goals
 - Publish a very detailed data flow diagram
(This system will need to be compliant with the larger ADFG data management systems)
- Next step is to build a prototype information system...that way they can get feedback from the users... Answer questions like: what added functionality are they looking for?
- Then need to isolate user cases. Test it out with a manager running the query, getting the data they need
 - Draft software requirements specification, functional and non-functional components

*Iterative realignment is key to the process; it acts as a feedback mechanism between users and developers.

Prototype

Important to help users realize the tech capabilities that are out there. Start by providing them with concrete examples, not necessarily salmon-specific.

Preliminary design considerations

- Make things web-based as much as possible.
- Make things fast, transparent, transferable.
- Georeference historic data as much as possible. Spatial data is a key component of the entry application from here on out. (linked to AK's anadromous waters catalog- AWC)
- Centralized storage platform for all data (tabular and spatial): SQL Server 08
- Data Structures for historical data will be modeled after existing ADFG databases.

Example System (LGL Chuitna: Mark-Recapture ASL study)

Questions for Steve Moffitt from facilitator:

- Summary of where you are at? Essentially just beginning. Still want to figure where we are going.
 - What are obstacles? Personnel and availability
- How much change in protocols and procedures will data management/dev encounter in the future... The switch to DIDSON sonar for escapement, create the need to change escapements to match older data. Most of other escapement data has been collected the same for last 40 years.

2. Department of Fisheries and Oceans, Canada (DFO) - Brian Riddell

- Talks about the "3 BIGS" of salmon conservation in B.C,
 - Need a full reorganization of DFO datasets.
- Goal: pertains to Canada's new Wild Salmon Policy (WSP). Very ambitious program on how we manage wild salmon.*
- Break out catch data and relate them to species-specific conservation units (CUs)

*Currently, a huge task is communication. (i.e. If people want escapement data need to talk with Erik Grundmann., if want catch data need to talk to __, etc)
-Want a web tool where people can access all this data easily.

(Riddell, Beamish) NPAFC Goals: 5 nations coming together so we can assess all salmon populations.

Conservation Units

Why was the concept of “stock” not sufficient?

-Created as management units, but don’t actually preserve biodiversity; the composition within them was also not fully understood.
-stock does not address the underlying processes among populations (adaptability)
-Guiding principles of the WSP: Conservation of wild salmon and their habitat is the highest priority. [Strategies (1-6)]

-CUs just being wrapped up in the Yukon.
~8200 spawning locations in BC and Yukon organized into CUs.

Methods for Defining CU spatially

32 Freshwater Adaptive Zones

13 Marine Adaptive Zones...these are overlaid to make Joint Adaptive Zones. (JAZ)
(purely ecological-based)

Then based on salmon biological characteristics the JAZs are further lumped or refined.

WSP calls to define benchmarks for these conservation units. “biological status zones”

3 fundamental changes to past assessments:

1. definition of a “conservation” or Lower Benchmark
2. explicit consideration of diversity within CUs during assessment
3. Incorporation of precaution (buffers) to account for uncertainty in benchmarks

Where API fits in

Strategy 1 of WSP, Step3: Monitor and assess

Indicator systems, intensive monitoring, and extensive monitoring

Question from facilitator: Obstacles you perceive?

Getting anything out of Holtby. (CUs specifically) So far no strong negative feedback on the CUs.

Still need to do more to incorporate steelhead into this conservation plan.

LGL Limited - Karl English

Recent salmon data inventory and stock assessment reviews in BC

...2005 Initiated Data Inventory for Ecotrust. And Evaluation of the quality of datasets

2006 Completed data inventory for Salmon and Steelhead and Core stock assessments

2007 Salmon workshop to rate key streams

2008 Most recently linking streams to CUs

They have evaluated information for each index stream... field method quality rankings 1-5. poor (presence/absence) to excellent. (an estimate of high res from an unbreached fence count.)

Escapement monitoring summary:

They've cross-linked the CUs and Index Streams, and then are placing them in bins of monitoring quality. For sockeye 142 CUs are missed altogether. (of 230). Majority in quality 2-3.

Assessing stock and fishery status:

Combine escapement data with catch data for CUs

Next steps

- Complete review of index streams. One more round of evaluation still needed.
- Link index stream metadata to nuSEDS database.
- Review methods used to estimate total annual stock exploitation rates or each CU
- Generate abundance and exploitation rate estimates for each CU.

Obstacles? We've worked productively with all regions in BC, but some areas are easier than others. They see the value in the work. And others find it more of a hassle. Key: Need to make sure everyone can see the end products. Early buy-in is critical. That will lead to better contribution.

3. Oregon Department of Fish & Wildlife (ODFW) – Jeff Rodgers

Web-based data access tool

Current state of the data management for ODFW recovery planning.

Why we've failed:

1. lack of resources. It seems more dire the state the less resources and more frantic the management becomes.
2. trying to solve everything once

Why we will succeed:

1. We are starting with a tractable problem.

Start with information needed to track progress of Oregon Coastal Coho Conservation Plan. Use this as a template for other ESUs in the state.

Another reason for choosing Oregon Coastal Coho is because it's been the most solid and ambitious monitoring program since 1998.

4 Major monitoring programs

- in-stream and riparian habitat, water quality
- Juvenile abundance and distribution,

- marine and freshwater survival
- Spawner abundance and distribution
- *Very data-rich environment

OR Coastal Coho Conservation Plan
Measurable criteria to meet desired status
Independent and dependent pops

Project background and rationale

- Individual projects have pretty good database systems but not well integrated. Takes too long for staff to retrieve data.--> recovery staff need ALL available info
- Recovery staffs do not have people or time to respond to data requests
- Misuse of data (staff turnover; don't know what data are in house)

Project vision

Readily available info for all users
Less time and resources required to respond to data requests
Less misuse/misunderstanding of data (descriptors of individual projects)

How can colleagues help you make progress on this project??

Provide us (ODFW) with ideas, past experiences on website organization.
Discussion on how this product fits with regional data management efforts.
*Where are we? Constraints? Lack of funding and lack time. *Not lack of data*

Who's doing what – solutions and specifics

Presentations from participants who are in the process of or have completed similar work.

A. Ecoinformatics at NCEAS – Jim Regetz (NCEAS)

- NCEAS Mission: address big questions in ecology by fostering collaboration among diverse experts. integrate existing environmental data.
- EcoInformatics goal: facilitate scientific community's access and easy interpretation of data.

Ecoinformatics program

- Research: better managing heterogeneous data, interpreting data using semantic technologies, capturing scientific workflows,
- Tools and services: metadata standard (EML), online repository (KNCB), advanced storage system for data/metadata (Metacat), Morpho, Kepler

Focus was on metadata early on.

-Makes data useful. Used EML, ecological metadata language, as a standard. Developed by scientists, for scientists. Easily formatted into html page for example.
How do you create? Simple web form is first option. More advanced option is Morpho desktop software. Morpho also has strong search capabilities to locate metadata, by keyword, query, etc.

Knowledge Network for Biocomplexity. (KNB)

Distributed set of repositories. Data maintained by individuals, access controlled by individuals, immediate change tracking, redundant backup. *NCEAS is not actually doing any data management themselves.

Participants: nceas, lter, obfs, pisco, esa, ucnrs

What NCEAS is doing now...

Going from documentation of data to its interpretation.

Need something more to capture the actual interpretation itself.

Semantic annotation

-mapping metadata and data contents to the rich set of terms and concepts that we understand as scientists.

-Lots of potential: navigation of relationships, include related terms/concepts, provide alternative interpretations of a dataset,...

-Semantic enhanced search. (e.g. if you search for Santa Barbara you still get info about CA... **Understands similar search terms)

-Enhanced Data interpretation (interprets the intent of fields as opposed just their definition themselves) and data interoperability.

-Data is connected to an Ontology through structural metadata and semantic annotation.

From data to analysis

Kepler workflow software. (OS) Create workflows to run analyses. Can use code from R, or MATLAB for example.

Applications of this work in the near future??

First steps, Identify high quality production data sets, and use EML for documenting these data. Need to join forces, organize biologists to start working on these ontologies.

May not have early returns, but will have long term value. Ontology will help get people onto the same page.

What you would plan to do next?

Start with some small dataset and see if you can summarize the dataset in this way using these semantics technologies.

B. Concepts and Tools for Fisheries & Watershed Analysis – Art Tautz (BC Ministry of Water, Land, and Air Protections)

Key questions

What are the ecological units for analysis?
Should we enhance existing information systems or develop our own?
Use OS or commercial, proprietary software?
Raster or vector?
Are there adequate resources available?

Components of info systems. Concentrate on storage, analysis and visualization

BC Provincial Systems

-Application (fisheries watershed atlas, land and resource data warehouse, image warehouse, habitat), storage (dbIV, oracle, analysis/formats, viz (2d,3d, stereo, animation...))

BC watershed atlas

Organized by Watershed code or (Pcode). 1:50k res. 3rd order watersheds as a basic unit. Watershed statistics and other environmental characteristics are attachable to the atlas
*Lost of applications of the atlas: hydro modeling, ecoregional planning, risk assessment, fish distribution, stream carrying capacity, salmon steelhead CU, cumulative impacts, knowledge management, climate change modeling.

Looking in to Raster based tools

-Example: "Hectares BC"; Start attaching attributes to 1 hectare grid across the province. No need for spatial joins all stored in a database. Very fast to run queries.
-Bulkley 3d raster. Viewshed models. Getting closer to what the real world would really look like.

Animation example: Want to visualization real time what the ocean would look like.
Answer questions like: how would this vary under different policy scenarios?
Terrestrial viz. connect to hectares database.

C. AYK Database Management System – Holly Krenz (ADF&G)

Arctic-Yukon-Kuskokwim regional background.

*Huge geographic scope. 150 projects that monitor salmon. Salmon observations on more than 315 streams. 70K records of daily/annual estimates of abundance

Data issues

Issues with data loss, different formats, and versions of data
No standards codes or definition used from area to area. Undocumented changes in collection methods.

Project Goal: Provide managers, researchers with better data access to useful information.

Project History: Surveyed users for dataset priorities, data rescue (thousands of ASCII files), aggregation of data into MS Access dbs.

Current project state

- Centralized AYK database with a web service to extract necessary info (within the AYK DB there is a metadata db with sub-repositories: ASL, test fisheries, escapement, etc...)
- Metadatabase: Descriptive data for all AYK projects, used to navigate among datasets
- Search Form to search for datasets
- If you are logged in...you can go in and edit, delete metadata; that way project managers can go in and make changes.
- Desktop Client: .Net clickonce app, Data import and export tool.

Issues encountered

Staff turnover. Disparate data sets. Multiple funding sources. New technologies. Time. historical data issues (making it fit the current standards). What defines a “project” (definition has changed a few times). Location (gis solution)

Next steps

Updated database with current years of data. Continue data cleanup. Improve data importer for client. Create reports (web app, client)

Future Goals

In season escapement counts. Better in season tools, incorporate other data,...

D. The Upper Columbia Approach to Monitoring and Data Stewardship. – James White (Upper Columbia Salmon Recovery Board)

Habitat Adaptive Management Framework

Adaptive management: More successful as an idea than as a practice.

Need to start with a question. End with the question answered

Question asked->Information needs-> monitoring objectives-> monitoring methods->monitoring-> data flow/collection-> analysis-> answer the question

*We do a lot of data collection that doesn't necessarily answer a question. not good use of time/money

Key Points to the UC approach

Identification of key management questions.

Data gaps prioritization (identify gaps by different tiers, 1-4)

Data inventory.

Local technical advice

ISEMP and other data management tools

*Upper Columbia Monitoring Strategy.(key document) Has plans specific to each sub basin. (e.g. Wenatchee)

Importing ISEMP work into a standardized database.

Protocol documentation
Regional Technical Team as technical advisors

Scalable Organization

- ISEMP program: all different survey types are all integrated together.
- Okanogan Basin Monitoring and Effectiveness Project
- Methow Restoration Council
- *All 3 need to be incorporated together into the Upper Columbia Monitoring Data Flow.

Draft qa/qc framework. Need to be applied at multiple levels. Not just the work of the data steward.

Habitat Work Schedule (public portal)

WDFW w/ interlocking software
A way for tracking what you are doing to see how it affects populations

E. ISEMP Data Management Approach – Steve Rentmeester (ISEMP/PNAMP)

Integrated Status and Effectiveness Monitoring Program (ISEMP)

- status and trend monitoring and pilot effectiveness
- design and run watershed-scale experiments (IMW)
- develop, test, implement (sample designs, data tools, eval tools, metrics, indicators...)

Work Flow

Questions to answers via regional sharing of the information.

Current state of work flow vs. future work flow

Current issues in flow

1. Metadata does not exist in most cases, currently.
2. Data entry is not validated.
3. exchange formats have not been defined (vocab is not consistent among organizations.)

Goal

Data Sharing (short term) and Data Interoperability (long term goal)

Tasks to correct work flow

Best Practices document is in development, will help advance towards data sharing

Protocol manager/protocol library. Storing info about protocols

Data entry application

Monitoring terminology glossary

Data portal (data sharing tools)

XML data exchange

ISEMP DB Management System

Includes wide range of survey types with different protocols

Data Model

Data Collections Events (made up of project, site, protocol, stat design)
These lead to “observations.” (e.g. spawning ground data model)
The data model describes “what is observed”

ATM: automated template module

Form for entering data collection events, metadata, observations, etc...
ATM form is updating on the fly as to what attributes show up in the form (depending on the protocol). Answer what data do I need to have the user enter?

F. Salmon Population Status & Trends – Scott Rumsey (NOAA)

Task 2 Abundance and Productivity

What are the benefits?

*Fill data gaps, increased certainty, minimize redundancies, and improve cost effectiveness. (state, federal, salmon recovery regions would benefit)

Task 2 Work plan

Data Integration and Sharing. Collaborative regional monitoring strategy.
What data is being collected? What data do we need?

Creation of a Data dictionary.

Data->Method->Metric->Method-> Primary Indicator

A lot of this, or similar work, has been done before, (e.g. State of the Salmon inventory) but doesn't indicate certain info like whether it is a population or productivity estimate.

What's left to do

- Still need to review and finalize the draft data dictionary.
- Work to develop an inventory tool
- In early 2009 will conduct a pilot inventory. With a data rich area like OR coast Coho, And then do initial gap assessment. Where are certain populations lacking data completely? Want it to be collaborative process with states, tribes, etc...

NOAA is prepared to throw resources and staff towards the pilot inventory. Beyond that they may need to look to BPA to fund similar project throughout the region.

G. SPS and GAPS Applications – Jeff Cowen (NOAA)

Technology

Oracle 10G – backend database.
Oracle Application Express (rapid app development tool)
Google Maps API

Salmonid Population Summary (SPS)

- created to meet NOAA ESA listing requirements - every 5 years.
- TRT population summarized data - spawner abundance, hatchery fraction, harvest, age structure. Raw data not included in the system.

SPS database interface

- easy to use web interface. Query by recovery domain (esu, mpg, pop)
- export to salmon population analyzer format or excel (SPAZ: www.nwfsc.noaa.gov/trt/wlc/spaz.cfm)
- other capabilities for restricted users like data and reference editing, creation of a data archive.

Future Goals

- Link sps/spaz results to restoration projects (PCSRF, etc)
- enable web map app for population selection. Map display of population attributes. (% hatchery, avg. abundance...)
- make site public

GAPS Database

Genetic Analysis of Pacific Salmonids (funded by PSC)
Central repository for lab genetic data.

Goal for GAPS

develop a db application to support the dissemination and growth of the baseline
Currently only microsatellite data has been loaded. Next will be adding some of the SNP data.
Future goal is to standardize the data as well.

Day 2 (11/4/08)

Who's doing what – solutions and specifics (continued)

Presentations from participants who are in the process of or have completed similar work.

H. Monitoring Glossary – Russell Scranton (PNAMP)

PNAMP Project to help define terminology

Goals

- create succinct list of indicators and metrics for monitoring
- provide keyword thesaurus
- reconcile differences in terminology between monitoring groups

*Based on an environmental data schema. (performance measures, indicators, metrics,...)

*Linked to much existing work (pnamp inventory , “Task 2” data dictionary, ISEMP...)

Next steps

- want web-based data entry capability.
- Will be having people adding content, and then validating.
- have panel of experts review and approve standards

Who would end up using the data entry tool?

People use this for use of lookup tables that people might need for creating forms

I. Pacific Northwest Aquatic Monitoring Partnership – Sean Quigley (USGS)

Goal of PNAMP

collaborative forum. Bridge gap between data practitioners and decision makers
Be a clearing house for certain data types

Current emphasis

Data management workgroup. (forum)

- Development of a library of long term aquatic monitoring protocols
- obtain knowledge, needs, and lessons learned through tech transfer meetings

Data management workgroup: tool dev, metadata, data storage delivery, data sharing
option for doing crosswalks (XML schema) – get data talking to each other.

Protocol Library

- metadata to doc protocols. Library provides that metadata for various tools. Should act in a “prescriptive” manner for new data and “descriptive” for historical data.
- Ability to exchange information between different groups more easily.
- Hope to influence future data collection efforts
- Can use the metadata library to find the appropriate methods for measuring something.

Current status of protocol library

completed the evaluation of protocol manager and protocol library recommendation

Next steps

- “now or never” db decisions. (metadata)
- Right now it is just a desktop app. Want to make it web enabled. And then put together a focus group of users- meetings to test it out.

Technology transfer meetings

- Share lessons learned from similar projects. Answer questions like what would you do differently
- Solicit review of existing tools.

Software platforms for protocol manager?

Vb.net works well, but searching capabilities are a bit weak.

Database is MS sql server. Don't know where the server is going to exist right now.

FACILITATED EXERCISE

How to not reinvent the wheel

Low Numbers: weak sustaining/restraining factors

High Numbers: strong sustaining restraining factors

RESTRAINING

1	Separate development within a single agency of the same data product – lack of communication
2	Provincial (not necessarily in the) Outlook (non-interconnectivity, etc.) – building regional or state or national tools and systems using closed or proprietary software and without regard to national/international efforts and standards.
3	Lack of metadata or support to generate it
4	Data documentation (metadata) <ul style="list-style-type: none"> • not enough time (“putting out fires”) • no requirements, not institutionalized
5	Too many pilots, not enough implementation
6	Silos - Departmental, political boundaries often necessary but often short sighted
7	Complexity <ul style="list-style-type: none"> • schemas • formats • standards too complex they won’t use it
8	Talking about specific tools before identifying needs
9	Disconnect between regional data summary needs and large centralized databases
10	Everyone supports standards as long as there is the standard
11	<ul style="list-style-type: none"> • Good idea • Initial energy/work • Lack of follow through
12	Lack of data product ownership
13	Designing for single protocol or organization
14	Lack of auditing interpretation of data standards plus definitions
15	No Buy-in from users
16	Metadata last – by wait to do metadata until after all of the data has been analyzed. Metadata becomes all pain and little benefit.
17	<ul style="list-style-type: none"> • Work done in Silos • Lack of communication results in redundant effort
18	Design by: <ul style="list-style-type: none"> • buzzword • mission statement • etc.
19	Failure to link data management to emerging key questions
20	Trying to design an all encompassing data management system that becomes unwieldy, inflexible, if outdated
21	Imposing a one size/type/style design/mentality
22	Priority for data management varies greatly among participants, both in general need and focus
23	Outline data libraries;

	<ul style="list-style-type: none"> everyone needs to use them need adequate search function and metadata
24	Lack of or irreconcilable location info on monitoring activities and population units
25	Failure of past efforts discourages buy-in to new ones
26	Sig. investment of time and expertise in existing systems – resistance to change
27	Top-down organization with rigid standards and centralized control
28	Coding too complex to react timely to changing needs
29	Time it takes to find data if not readily accessible
30	No consultation with end users, so no buy in or use of final product
31	Not enough consideration of what it will take to maintain the system nor the burden it will place on those responsible for putting data in
32	Not enough direct client and developer <u>communication</u> – supervisor can be unnecessary middle man
33	Planning for technology, forgetting about people
34	Database or tool development that is NOT informed by management questions
35	Institutional commitment (could be \$ or a behavioral expectation) to preserving info and knowledge ABSENT
36	Regional managers housing data (i.e. escapement) instead of uploading to NUSEDs (gov. db)
37	<ul style="list-style-type: none"> Lack of flexibility in reference to emerging technologies System can't evolve Compartmentalized architecture
38	Lack of resources (staff, funding, time, etc.)
39	Most failed or incomplete efforts; <ol style="list-style-type: none"> lack of adequate funding staff turnover
40	Ignoring end user/client experience
41	<ul style="list-style-type: none"> Not an agency priority Not recognized as important to decision makers Ex: GIS coordination
42	Lack of or inconsistent documentation about the appropriate <u>uses</u> of available data (area of inference, quality, etc.)
43	Lack of understanding by policy makers and leadership of need for resources to manage data and importance of data management
44	Lack of Buy-in from upper level managers (decision-makers) contributes to lack of funding
45	Lack of adequate funding
46	Systems fail to meet the needs of the data producers
47	Lack funding and other resources to do the job that's needed
48	Insufficient commitment of workforce
49	Failure to achieve Buy-in at the operational level – eg. A metadata system is useless if no one commits to creating the metadata as agreed upon
50	Lack of consistent regional direction

SUSTAINING

1	Automated email notification of updates
2	Cross – pollination and collaborative efforts
3	Data sharing helps open lines of communication
4	Prototype early
5	Support from upper management
6	Specific and general user functionality
7	Component based architecture – loosely coupled pieces: reusability ↑ and cost of change ↓
8	Flexible and expandable

9	New system improves and mirrors prior system
10	Online data libraries; <ul style="list-style-type: none"> • facilitates collaboration • new post or update notification • data file sharing
11	Transferability/reusability; <ul style="list-style-type: none"> • open source lock/design • documentation • cross platform • deployable
12	Generic data model to allow variation across use groups and over time
13	Maintains data neutrality – leaves analyses to the experts
14	Database mgt expertise in regional data collection centers
15	Flexible structures and distributed data
16	Whenever possible choose the more flexible alternative
17	NCEAS tools as Prototypes– EML, Morpho, KNB, Kepler,
18	References all data sets/maximizes transparency
19	Open source code; <ul style="list-style-type: none"> • No black box code • Code sharing • Saves time and \$
20	Developing tools that facilitate access to important data (catch, escapement, harvest rate)
21	Open source
22	Allows data producers to manage/maintain data in native format
23	“freeware” to facilitate/encourage participation by local entities in data integration and sharing
24	Ontologies
25	Open standards and open source software
26	Founding entities impose consistent data management requirements; <ul style="list-style-type: none"> - data dictionary - data sharing - etc. - schema
27	Standardize language
28	Evaluating data quality separating the good from the bad
29	Frequent communication among group development members
30	Needs/critical requirements analysis – get the carrot before the horse before the cat...
31	Designing from broad perspective
32	Separate storage and analysis applications
33	Sustained \$ or expertise (resources) to keep products “fresh”/useful/valuable to users...because use needs and expectations Δ
34	Clear definition of end products + key data needs
35	Avoid one-off solutions (you are not alone)
36	Provides funding to participate
37	Spending a lot of time to come into agreement on common standards making sure all users well represented
38	Question-based design
39	Data steward
40	Metadata first! – document metadata early in the workflow. By documenting during the strategic planning and protocol development phase, we can then use the metadata to support the rest of the workflow.
41	Modular or scalable/extensible systems that allow easy Δ or modifications to reflect Δ needs

Discussion – how to not reinvent the wheel

How do we target our audience? What stands out?

OPEN SOURCE

- We all want OS but sometimes we just can't go there (bureaucratic reasons)
- Key to understand the difference between OS code and software. We can still make code publicly available.
- There are even challenges there where some organizations can't expose their code (e.g. ODFW)
- Need to have executive buy-in or money to create incentives to change behaviors.
- Do agencies have the expertise to maintain OS systems even if someone is able to create it for them?

THE BROADER CONNECTION BETWEEN PROJECTS ACROSS BOUNDARIES

- Building on the successes of previous database systems that have worked. (e.g. SPAZ)
- You cant have the end product without the core datasets. Incentive to maintain the core datasets.
- Strategic Documents are important (Hilman Upper Col) what is the relationship between the raw data and indicators? To figure out what the important metrics are.
- Proof of concept to funders/decision-makers; make things transferable and transparent, key to API projects.

USER FEEDBACK IS KEY

- Everyone wants to be accommodating to all the data needs.
 - Leadership doesn't know there is a better way to the current data system because they haven't seen that there is "another way to live"
 - They need to see the potential, and once they do, it will be easier to break that barriers.
 - Needs to be flexible and easy to use.
- Useful to separate user needs and wants. Some changes we need to make and user may need to be flexible in some regards.

COMMITMENT TO COORDINATE FOR THE LONG-TERM

Key even when we disagree. Need to stay committed. Need to make sure people don't just walk away. (maintain resources: staffing and funding)

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.

BEYOND THIS MEETING

Next steps

Cathy

SOS needs to stay in touch with PNAMP to help move along some of their agenda. Bring in the AK and BC folks into the data dictionary discussions, to help move the timeline along.

Still have fundraising to do.

Karl

Not much overlap with groups in the short term. But as we head more into the output phase need to start talking with folks who have already made these steps.

Jeff

Going to schedule a meeting to bring PNAMP, ODFW, PSFMC (Streamnet), NOAA, Ecotrust folks together to nail down the data conduit.

Erin

Julie and Erin will be meeting with developers to go over specific use cases to define deliverables

Steve

Still need to do the requirements analysis

Start to inventory which data we want to salvage

Start a data dictionary, see if we can use preexisting

Create metadata for historical data they want to salvage

Tim

Reviewing everything that has come of this meeting. Looking at PowerPoints and meeting notes. Lots information, but wasn't really enough time process it all.

To do list...

- add PowerPoint presentations to SalDAWG Central Desktop site

-Could participants please post materials/publications (from their presentations) to Central Desktop.

→Scott Rumsey – to post a copy of white paper on SPAZ

→Jim Regetz – to post something on Kepler

-It would be good to have a checklist synthesis of the 3 different projects. (Cathy with project help)

-re-assess program on key sustaining/restraining factors

-effectiveness to-do (e.g. IRC, Central Desktop = simple contact list using extranet profiles)

FEEDBACK

SalDAWG Future Meetings and Topics

Next on the Agenda?

Meeting at ODFW in the early summer 2009. Data Visualization will be the focus;
Exploring innovative means for displaying/conveying natural resource data.

For next meeting, would like to talk to decision makers, GBMF, funders, to see what they are anticipating to see as outputs. That way we can help deliver those products. (Karl)

- e.g. AK requirements analysis doc will be addressing some of this issue,
- e.g. executive NWEIS high level performance monitoring;
- e.g. Columbia River Power data deliverable

What place does SalDAWG have among all the other data improvement groups, meetings, gatherings? (i.e. what makes this collaboration unique; how can we add value)

- This group is more about pilot projects and really demonstrating some of the things going on.
- Demonstrates some of the added value of these tools to the rest of the state/region etc.
- There are actual management products that are the outcome of the API (not just a discussion)

What aspects of the meeting worked well? (+)

- informal interactions (beer+++)
- planners orienting presenters
- plan “next steps” with deliverable and deadlines
- diversity of participants

What aspects of the meeting might you change (Δ)

- more pre-meeting clarity for designed meeting outcomes and key people to participate
 - summarizes meeting support info and about meeting outcomes
 - try to create more work time – less presentation time
 - executive summary packet for airplane read
 - presenters prepare clear, concise, summary of key points
 - do not redo what we did at this meeting
 - original content
- Better time enforcement for presenters
- work with topical colleagues
 - work with different professionals
-

Attendee List | SalDAWG Meeting 1: Nanaimo, BC (November 3-4, 2008)

First Name	Last Name	Organization	Email
Anita	Blakley	LGL Limited	ablakley@lgl.com
Robert	Bochenek	Axiom Consulting & Design	rob@axiomalaska.com
Cedric	Cooney	Oregon Department of Fish and Wildlife	cedric.cooney@oregonstate.edu
Jeff	Cowen	NOAA Fisheries	jeff.cowen@noaa.gov
Karl	English	LGL Limited	kenglish@lgl.com
Julie	Firman	Oregon Department of Fish and Wildlife	julie.firman@oregonstate.edu
Erin	Gilbert	Oregon Department of Fish and Wildlife	erin.gilbert@oregonstate.edu
Erik	Grundmann	Department of Fisheries and Oceans Canada	grundmanne@pac.dfo-mpo.gc.ca
Cathy	Kellon	State of the Salmon	cathy@ecotrust.org
Heath	Kimball	Alaska Department of Fish and Game	heath.kimball@alaska.gov
Holly	Krenz	Alaska Department of Fish and Game	holly.krenz@alaska.gov
Rich	Lincoln	State of the Salmon	rlincoln@wildsalmoncenter.org
Stephen	Lloyd	Ecotrust, State of the Salmon	slloyd@ecotrust.org
Steve	Moffitt	Alaska Department of Fish and Game	steve_moffitt@fishgame.state.ak.us
Bruce	Patten	Department of Fisheries and Oceans Canada	pattenb@pac.dfo-mpo.gc.ca
Sean	Quigley	USGS/PNAMP	squigley@usgs.gov
James	Regetz	National Center for Ecological Analysis & Synthesis	regetz@nceas.ucsb.edu
Steve	Rentmeester	ISEMP/PNAMP	steven.rentmeester@noaa.gov
Brian	Riddell	Department of Fisheries and Oceans Canada	brian.e.riddell@dfo-mpo.gc.ca
Greg	Robillard	consultant	gxavier@gmail.com
Jeff	Rodgers	Oregon Department of Fish and Wildlife	jeff.rodgers@oregonstate.edu
Scott	Rumsey	NOAA Fisheries	scott.rumsey@noaa.gov
Russell	Scranton	NOAA Fisheries	scranton@noaa.gov
Art	Tautz	Ministry of Water, Land, and Air Protection	art.tautz@gov.bc.ca
Tim	Welch	State of the Salmon	twelch@ecotrust.org
James	White	Upper Columbia Salmon Recovery Board	james.white@ucsrbc.com