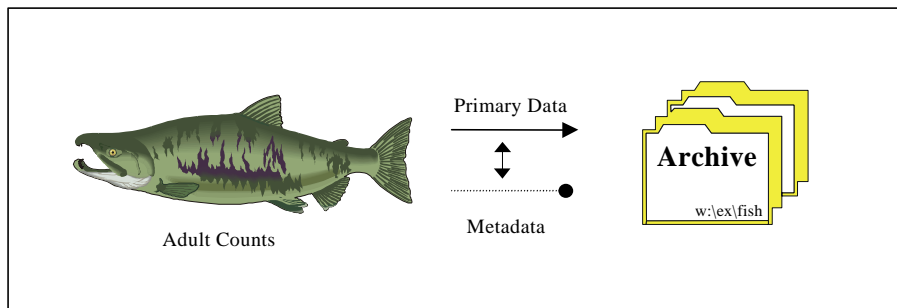


# Independent Scientific Review Panel for the Northwest Power Planning Council

## Review of Databases Funded through the Columbia River Basin Fish and Wildlife Program



Charles C. Coutant  
Daniel Goodman  
Susan S. Hanna  
Nancy Huntly  
Dennis Lettenmaier  
Lyman McDonald  
Brian Riddell  
William Smoker  
Richard R. Whitney  
Richard N. Williams  
Scott Urquhart, PRG

ISRP 2000-3  
May 11, 2000

# **Review of Databases Funded through the Columbia River Basin Fish and Wildlife Program**

**Contents:**

ASSIGNMENT, SUMMARY AND RECOMMENDATIONS.....	1
PROJECTS REVIEWED.....	5
INTRODUCTION .....	5
DATA EXISTENCE.....	6
Data Definitions .....	7
Data Needs .....	8
Standardization of Methods .....	10
Sampling, Evaluation, and Inference .....	11
Data Ownership .....	11
DATA ARCHIVING .....	12
DATA ACCESS .....	15
Distributed Access .....	19
Appendix 1. Assignment from the Northwest Power Planning Council .....	21
Appendix 2. Brief overview of projects reviewed, summary of ISRP briefings, and documents presented to the ISRP for review .....	23

# **Review of Databases Funded through the Columbia River Basin Fish and Wildlife Program**

"Where is the knowledge in all the information?  
Where is the wisdom in all the knowledge?"

T.S. Elliot

## **ASSIGNMENT, SUMMARY AND RECOMMENDATIONS**

The ISRP's past two annual reports (ISRP 98-1, ISRP 99-2) have included recommendations for an independent review of databases that are funded through the Columbia River Basin Fish and Wildlife Program (Appendix 1). There are five data management projects and others in the planning stages. The ISRP recommended that this review take place before funding is continued beyond fiscal year 2000. The Council approved a review of database projects and specifically asked the ISRP to address the following questions (paraphrased from Appendix 2):

1. Do these projects supply the region with information **NEEDED** to plan and manage the Columbia River ecosystem? What are their strengths and weaknesses? How do they address the needs of the diverse audience including researchers, managers and decision-makers?
2. What type of data structure is needed to ensure that the necessary data are accessible for analyses to inform decision making in the basin?

In addition, as PATH transitions out, how should the Council ensure the continuation and improvement of structure analytical functions such as collaborative analysis, maintenance of derived data, data mining and validation.

### **Short Answers to the Questions**

- The first set of direct questions is easy to answer, in part because of the emphasis we have added by capitalizing the word "NEEDED." The short answer to this question is simply "No." The database projects do not supply the region with the information needed to plan and manage the Columbia River ecosystem. However, it is extremely important to understand that this failure is not a failing of the database projects. It is, first and foremost, a failure of the region to collect the needed data. Many crucial quantities are not being measured at all, some are measured haphazardly and in inconsistent ways, and a few of the important quantities that are being measured do not get compiled in any of the central databases. It is only a minority of the "needed" information that is the subject of ongoing, systematic monitoring. The data management is being handled reasonably well for the resulting data that is sent to the projects.

- The answer to the second question concerning data structure is that the tasks of compiling, archiving and providing access to the data can best be handled by a decentralized distributed access system. This is the direction in which the various database projects are presently evolving.
- As PATH transitions out, we recommend that the Council provide strong directions to the program's data access sites (DART, part of the Fish Passage Center, and part of StreamNet) to ensure the continuation and improvement of analytical functions, maintenance of derived data, data validation, and service to data users. We would observe that PATH personnel are talented individuals who might be encouraged to respond to needs for expanded staff within the present data access sites, within private data access sites which are expanding in the region, or to Requests for Proposals released by the Council for new data access sites to meet specific needs.

### **Summary and Recommendations: Data Existence**

- No organization presently is taking responsibility for comprehensive design of data collection in the basin. This problem exists at two levels. At the first level, the data access centers find that their users frequently request data that are not being collected, or that the usability of the data is hindered by inconsistent protocols used in their collection, with the effect that the centers are limited in their ability to respond to user requests. As a first step to resolving this problem, the ISRP recommends formation of a joint working group of the existing data centers to develop recommendations for standardization of data collection protocols and to identify data deficiencies.
- At a broader level, the ISRP finds that the region has not developed forward-looking monitoring programs designed to address questions that may arise in the future. This problem goes well beyond the charge of the current data centers, and is central to the ability of organizations within the Columbia Basin to achieve their goals. As a first step in addressing this problem, the ISRP recommends solicitation of proposals for a comprehensive monitoring program designed to provide the data needed to meet the goals of the Council's forthcoming revision to the Fish and Wildlife Program (FWP). The ISRP expects that such an expanded monitoring program would draw heavily from existing efforts, but would need to go well beyond them to meet the central objectives and goals of the revised FWP. Some data needs observed by the ISRP over the last three years are listed to initiate discussion on establishment of an expanded monitoring program.
- The basin states (Idaho, Montana, Oregon and Washington) and other interested parties, such as the tribes and federal agencies, need to standardize their methods for collecting data. The Council should play a top-down role working through the respective Governor's Offices, State Departments of Fisheries and Wildlife, and the funding of projects in the Council's Fish and Wildlife Program to convince managers that personnel time should be allocated to this task.

**Summary and Recommendations: Data Archiving**

- The data archiving function should be modularized so that there is one data archiving center responsible for each class of data. Funding for data archiving should be on a project basis, where each project archives a particular class of data.
- It is critical that metadata be archived in a database structure that maintains the association between primary data and their pertinent metadata.
- We recommend adoption of a policy requiring that the reporting requirements for projects funded by the program include requirements for delivery of primary data, and their associated metadata, in a standard machine readable format, within a specified period of time. Compliance with this policy should be a condition for continued funding.

**Summary and Recommendations: Data Access**

- The data access function and calculation of derived data could be modularized in order to serve specified data access needs for purposes of data analysis. The network of modularized data access sites could evolve with changing needs for data analysis. The articulation of need should arise from the community of data users.
- It may be convenient, but it is not absolutely necessary to archive derived data, because with adequate documentation they can always be reconstituted from the archived primary data and metadata. Indeed, it is best that any archiving of derived numbers be flagged, so that users do not mistake them for primary data. It is worthwhile to revisit the calculation of derived quantities, at intervals, since assumptions may be revised with experience, and modeling and analytical methods may evolve.
- A portion of funding for data access systems could be self-generated by contracting with data users. This approach is currently being used by DART and some private data access centers. User fees (beyond a basic unit of courtesy assistance) might be charged by the Fish Passage Center and StreamNet to further augment their budgets.
- Competition might be encouraged between data access sites (DART, part of the Fish Passage Center, part of StreamNet, and private contractors) to provide data assembly and analysis services to data users. Data access sites should establish limited analytical capabilities to deal with questions from the public and government agencies. In short, the data access centers (particularly Streamnet) should be turned loose to do what they should do best, search, sort, and collate data -- possibly from multiple data archives -- for purpose of analysis by data users. This does not necessarily mean turning the data access sites loose on large-scale data analyses, but

rather emphasizing their role in accessing and assembling data that users can then analyze further.

### **Summary and Recommendations: Overall**

- There is no need to centralize the entire data storage and access system. Internet technology allows for a very effective distributed access system to be developed by web links between modular sites. These sites take responsibility for various functions, such as data archiving or data access, and may specialize in particular types of data, or particular types of analyses. Modularization and specialization among the various sites participating in the networked system should encourage efficiencies and allow for sound fiscal management.
- The amount of money that could be saved by curbing the small amount of redundancy that exists between data management projects (primarily between Fish Passage Center and DART) pales in comparison to the problem of meeting significant basinwide data gaps. We recommend continued support of the database programs. The Council focus should move on to identifying and filling gaps in data and metadata. The ISRP also supports the efforts of the Council's Columbia River Ecosystem Research, Monitoring and Evaluation data directory (the "yellow pages") and Bonneville Power Administration's Columbia River Fish and Wildlife Information Portal and encourages movement towards distributed access systems.
- A distributed access system is most easily implemented given standardized methods and metadata describing the information contained in each data set. The Columbia River Basin does not yet have standardized methods and this makes access to data more complex. Further, data archive sites do not currently provide all necessary metadata for their primary data. These are perhaps the biggest obstacles to implementing fully functional distributed access systems in the near term.
- The overall impression of the ISRP is that the various database centers are doing reasonably well with the data they are given. It is not their fault that data gaps exist or that incompatible methods are used by the various agencies in the basin. The centers could have done a better job in making metadata more readily available. However, we have the distinct impression that in some cases, their hands have been tied by policy decisions of agencies and oversight committees on both data and metadata availability. Those restrictions, most notably affecting Streamnet and reducing its utility to the region, should be removed.

## PROJECTS REVIEWED

The ISRP was briefed on the BPA-supported data management entities listed below. What constitutes discrete entities can be confusing, because each entity often has several names, and several discrete databases are maintained by some organizations.

1. University of Washington's Data Center, Data Access in Real Time (DART) [www.cqs.washington.edu/dart](http://www.cqs.washington.edu/dart); Second Tier Database Support for Ecosystem Focus (BPA # 9601900)
2. Fish Passage Center (BPA #9403300), [www.fpc.org](http://www.fpc.org); Comparative Survival Study (CSS) (BPA #8712702); Smolt Monitoring by Federal and Non-Federal Agencies (SMP) (BPA #8712700)
3. Streamnet (BPA #8810804), [www.streamnet.org](http://www.streamnet.org); see Pacific States Marine Fisheries Commission (PSMFC); [www.psmfc.org](http://www.psmfc.org).
4. PTAGIS (BPA #900800), [www.psmfc.org/pittag](http://www.psmfc.org/pittag); see Pacific States Marine Fisheries Commission; [www.psmfc.org](http://www.psmfc.org).
5. Coded Wire Tag (CWT, Regional Mark Information System (RMIS), BPA #8201300); see Pacific States Marine Fisheries Commission; [www.psmfc.org](http://www.psmfc.org).

In addition, the ISRP was briefed on the data efforts of the Ecosystem Diagnosis and Treatment project and two data access sites that are in development -- the Northwest Power Planning Council's Columbia River Ecosystem Research, Monitoring and Evaluation data directory ("yellow pages") and Bonneville Power Administration's Columbia River Fish and Wildlife Information Portal. More details on these ongoing and developing data management efforts are provided in Appendix 2.

## INTRODUCTION

In past reviews, the ISRP identified several concerns with basinwide database management. The ISRP found much potential for duplication of services, little justification for how an individual data system is distinct from the others in its scope or in the services it provides, a lack of coordination, and the potential for different versions of the "same" data, which could lead to confusion. Specifically, the ISRP reported that it could not find compelling reasons to fund multiple databases of smolt monitoring results.

During the current review process, it became clear that, with modern technologies, multiple data centers do not necessarily represent either duplication or inefficiency (though, of course, both are still possible). The real issue is functionality. With current technology it is feasible, and probably desirable, to separate functions among multiple sites, in a "modular" fashion. Thus, the overall goal of the review is to investigate whether the needed functions of a basinwide data system are well served with the present configuration, and to consider what changes might enhance the functionality.



There are three main types of function, namely:

1. Data existence
2. Data archiving
3. Data access

## **DATA EXISTENCE**

Data existence has to do with the overall design of data collection activities in the basin. Ideally, there would be a systemwide assessment of data needed to provide information for management decisions. Then some organization would take responsibility for designing monitoring and experimentation programs that could efficiently deliver those data. Finally, some organization would take responsibility for implementation of the design.

At the moment, there appears to be none of the components in place for dealing adequately with data existence. No overall assessment of specific data needs has been done, nor has there even been a systematic data inventory. Most data sets exist because of unique historical events and independent programs. No organization has taken responsibility for a coordinated basinwide design, and no organization has taken responsibility for uniform and consistent implementation of such a design. Indeed, it is not even clear whether any existing organization has broad enough authority to take command of basinwide implementation. Therefore, implementation might well require a new cooperative venture among several organizations. PATH, the Framework (including Ecosystem Diagnosis and Treatment), and NMFS's Cumulative Risk Initiative all have activities that bear on this coordination and implementation, but none have actually generated such an assessment and inventory.

We see several components of data existence: 1) definitions, 2) the need for certain types of data and metadata not now being collected, 3) standardization of methods, 4) sampling and inference, and 5) data ownership. We discuss these briefly below.

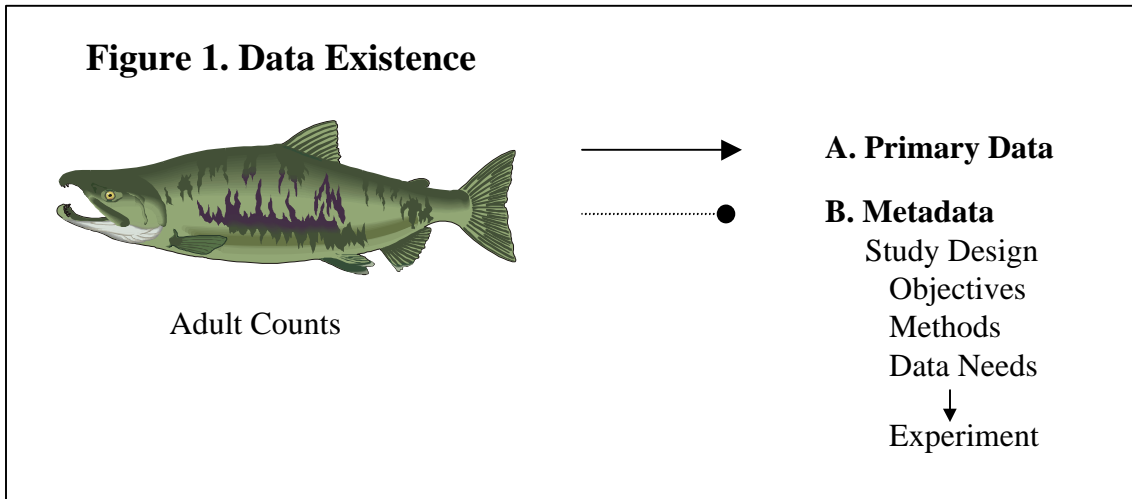


Figure 1. Data Existence. Data exist in two forms, as exemplified by adult fish counts. The primary data are the numbers of adult fish actually counted in a time interval and location. The metadata are the descriptors of how those numbers were obtained, including the monitoring design (selection of times and locations), objectives, and methods. Metadata would, for example, show that the counts were made for only 8 hours per day, perhaps necessitating subsequent extrapolation to 24-hour counts (which become derived quantities, see figure 3).

### **Data Definitions**

"Primary data" are direct measurements (Figure 1). They are not estimates, and they are not the results of expansion, conversion or adjustment. Examples of primary data would be: counts of redds on a defined section of stream at a specified time; the list of coded wire tag recoveries in a defined harvest landing lot; the weight of smolts released from a hatchery in a defined lot.

"Metadata" are the documentation of the study design: objectives, measurement methods, sampling design, and association of each primary data measurement with a time and place. The completeness and adequacy of the metadata are judged relative to the uses that we contemplate for the analysis and interpretation of the primary data. Ancillary information that is necessary for re-analysis and interpretation of data is "necessary" metadata.

"Derived quantities" are obtained by manipulating primary data to estimate a quantity of interest that is not directly observed. For example, an estimate of the number of redds in a watershed might be obtained by dividing the redd count (primary data) by the fraction of the suitable total stream length in the watershed represented by the sections of stream where counts were actually made (metadata). In this example, the assumption that simple division, based on linear stream measure is the correct "expansion" constitutes a "model" that may or may not be the best model. The documentation of how stream miles were

measured from maps (the result depends strongly on the resolution of the map), and how "suitability" of each stream section was determined constitute part of the necessary metadata.

QA/QC (quality assurance/quality control) information is the component of the metadata that documents the precision and accuracy of the primary data. This may include replicated measurements, calibration measurements of instrumentation, inter-observer comparisons, and verification that protocols were followed.

### **Data Needs**

Data needs are determined by the kinds of information needed for management decisions. Our concepts of what information are needed change as understanding is improved and some challenges are met. However, some types of data are fundamental. For example, the Council's goals for the Fish and Wildlife Program include numbers of adult anadromous fish and data on adult numbers are of the utmost importance. Of additional importance is information on factors that can affect the abundance of all life stages of anadromous fishes, leading to effects on adults. Similar needs exist in the resident fish and wildlife programs.

Inclusion of the following databases in an integrated network (with metadata) should be considered to provide decision-makers, the fisheries and wildlife community, and the public with the necessary information to interpret complex environmental relationships affecting anadromous fishes, resident fishes, and wildlife. The list below is intended to provide a basis to initiate discussion about what should be included as a minimum in a database system for Council decision making. Part of these are currently included in existing systems, although key elements with regard to the metadata may be missing.

A. Essential data for evaluating success of the Council's Fish and Wildlife Program in terms of adult salmon and steelhead:

1. Detailed harvest information (including coded wire tag data) by fishery, location, and time of harvest. Although not presently included, information on the specific location of capture could be added by using the Global Positioning System (GPS) for both ocean and freshwater harvest.
2. Fishing effort information by fishery, location, and time of harvest to cross-reference with harvest information.
3. Standardized escapement and reproduction data (e.g., redd counts, carcass counts, weir counts, juvenile counts) by mainstem location or stream location and date of surveys over years with metadata on selection criteria for survey sites within streams. Records should include subsequent extrapolated estimates calculated by participating organizations and metadata on methods used.

4. Indices of fish condition during upstream and downstream migrations (e.g., weight per length, disease, etc.).
- B. Data useful for monitoring and evaluation of the effectiveness of the region's fish and wildlife programs:
1. Hydropower-operations data for all mainstem dams on a unit-by-unit basis for the powerhouse discharge and spillbay-by-spillbay for spill volumes over time (i.e., hourly or finer). This information should also include forebay and tailrace elevations and head, along with total river flow.
  2. A GIS-driven database on watershed monitoring data should be established. Ideally, this database would be coordinated with terrestrial monitoring-mitigation data also being collected.
  3. All available stream flow and water quality information (especially temperature and dissolved gas) along with the dispersion of water monitoring stations.
  4. Information on hatchery operations for both resident and anadromous fishes: rearing conditions, production, release methods, and fish quality over time for all hatcheries. The Fish Passage Center provides some information on hatchery releases of anadromous juveniles. We heard during the review that BPA is working with an individual hatchery to make more than 10 years of data available in a database. The National Marine Fisheries Service is developing a template for completing Hatchery and Genetic Management Plans (HGMPs) for implementation of the Endangered Species Act and evaluation of other hatchery operations. The intent is to provide a thorough description of each hatchery operation, including the facilities used, methods employed to propagate and release fish, and measures of performance (<http://www.nwr.noaa.gov/1hgmp/hgmptmpl.htm>). These efforts should be standardized and expanded to all hatcheries.
  5. Data on tracking of radio-tagged fish.
  6. An inventory of past and current research within the Columbia Basin with provision for access to all reports. Public access to data and metadata should be a requirement of funding by the Fish and Wildlife Program.

C. Data judged to be insufficient during past ISRP reviews of projects and proposals.

Listed below are some data needs that currently are not being addressed in a comprehensive and systematic manner for the anadromous fish program. Similar lists can be provided for the resident fish and wildlife programs. This example list for anadromous fish is not prioritized and should not be considered exhaustive but is

provided to help generate needed regional discussion. Some items in the list have been measured in research projects but have not been measured for basinwide applicability.

- a) Growth and survival of juvenile fish while migrating from McNary to Bonneville, to the estuary and while in the estuary.
- b) Age-specific growth and survival in the ocean.
- c) Incidental mortality in ocean fisheries. .
- d) Adult survival in upstream migration to the spawning grounds.
- e) The epidemiology of BKD and possible transmission among fish.
- f) Competition between hatchery and wild fish throughout the life cycle.
- g) Predation of hatchery fish on wild smolts.
- h) Return rates from individual hatcheries.
- i) Straying of hatchery fish into wild populations.
- j) Reproduction of hatchery strays.
- k) Fitness of hatchery strays when they reproduce in the wild.
- l) Fitness effects of introgression of hatchery fish into wild stocks.
- m) Rates of selection when wild fish are taken for hatchery brood stock.
- n) Fecundity rates of wild fish.
- o) Egg-to-emergence and emergence-to-smolt survival rates of wild fish.
- p) The relation of the above life history parameters to habitat characteristics.

### **Standardization of Methods**

The basin states (Idaho, Montana, Oregon and Washington) and other interested parties, such as the tribes and federal agencies, need to standardize their methods for collecting and defining responses if the data are to be applicable for comparing subbasins and basin or province wide analyses. Classical measurements and methods must be maintained for a period of time concurrent with new standardized procedures in order to have a continuous history. After several years the previous measurements could then be dropped. The Environmental Protection Agency (EPA) regularly goes through such calibration exercises as new methods of chemical analysis emerge. Agencies argue that they must maintain their classical methods, but applicability of this continuity argument to EPA would mean that all chemical methods would be at least 30 years out of date.

The standardization can be accomplished by getting the affected parties to a workshop where the explicit charge is to standardize methods so the needs of all parties are met by common measurements. In the past, methods have been independently selected at different points in time by field personnel or others intimately involved with collection of the data. Appeals have been made for standardization of methods. Unfortunately, a bottom-up approach for standardization of methods has not served the region well. Field personnel frequently do not appreciate the importance of data standardization outside their location or topic of interest. The Council can play a top-down role working through the respective Governor's Offices, State Departments of Fisheries and Wildlife, and the funding of projects in the Council's Fish and Wildlife Program to convince managers that personnel time should be allocated to this task.

In the extreme case, ambiguous data collected by divergent methods may lead to decisions based strictly on political considerations, rather than with consideration of scientific information. We need to advance the case of comparable scientific data, so they can serve their appropriate role in decision making.

### **Sampling, Evaluation, and Inference**

Much of ecology, including fisheries science, has a tradition of hand-picking sites at which information is gathered. This is entirely appropriate for evaluating, for example, the effect of point source effects, such as spawning success following a habitat improvement project. Suitable responses should be monitored and evaluated at the project site and at appropriate reference areas. This allows unambiguous conclusions concerning the effect of the point source.

On the other hand many factors of interest concern large area or “landscape” effects. These include regional habitat degradation or improvement, and loss or colonization of species over broad areas. Inferences for such situations REQUIRE the use of probability samples (e.g., random or systematic samples) to represent the larger area. This will not be accomplished with sites chosen for convenience or political reason or by judgment. Probability samples can vary the sampling intensity according to interesting landscape factors and/or utilize stratification. Numerous examples exist that show how haphazardly gathered data, or even data gathered for another purpose, can mislead. For example, it is common knowledge that, until recently, the Oregon coastal coho survey sites were selected in some of the best habitat and trends on the selected sites did not predict the population trends (Scott Urquhart, Oregon State University, personal communication). A probability based sample of Oregon coastal streams has recently been implemented for coho surveys. Similar to the original Oregon coastal coho survey sites, the current sites for survey of chinook spawning in the John Day River are contiguous extensions of sites selected in the 1950s in known spawning habitat (ISRP 99-2). The method for selection of survey sites in all subbasins of the Columbia River, including the John Day subbasin, should be re-evaluated.

### **Data Ownership**

There has been a tradition, in many scientific disciplines, of individual investigators treating their data as "proprietary" until the analysis of the data has been published. The ostensible motivation has been assurance of priority of credit for original discoveries. In recent years, this tradition has been breaking down for a number of reasons: the reality of public funding creating an argument for public ownership (and in some instances access through the Freedom of Information Act), the frequency of collaborative research involving a number of investigators who share the data, the gradual recognition that in many areas of science huge stockpiles of data are accumulating faster than they can be analyzed, an awareness of the value of scientific syntheses of multiple related data sets

(in some fields this synthesis is itself a formal activity called "meta-analysis"), and new computer technologies that are making access and analysis practical. The result is the dawning emergence of a new "web culture" of open data access and open invitation to independent re-analysis and alternative-analysis. We consider this new culture a healthy development that will encourage quicker analysis, more powerful analyses, better documentation, and hopefully, easier scientific consensus.

All these considerations bear on the data generated with funding from the Columbia Basin Fish and Wildlife Program. We recommend adoption of a policy requiring that the reporting requirements for projects funded by the program include requirements for delivery of primary data, and their associated metadata, in a standard machine readable format, within a specified period of time. Compliance with this policy should be a condition for continued funding.

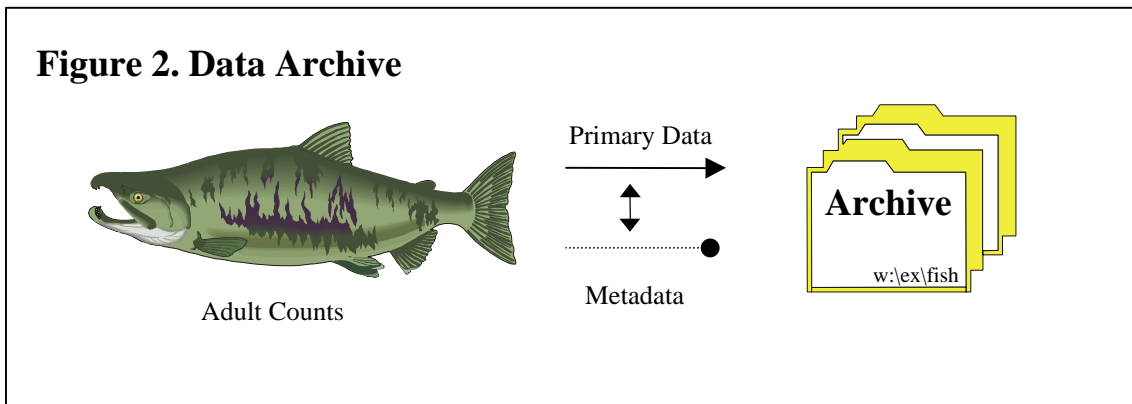


Figure 2. Data Archive. Primary data and their associated metadata are stored in a data archive. In a distributed data system (see following figures) the data archive has a unique Internet address through which its contents are accessed.

## DATA ARCHIVING

Data archiving is the compilation of primary, measured data and associated metadata (Figure 2). This is done in a single facility under a quality assurance/quality control (QA/QC) process with provisions for maintaining data integrity for some class of data. There must be consensus among all the parties that generate data of this class that this archive is the single, most appropriate archive. Use of these data, by the archivist or others, for analysis is done in a "read-only" mode: this does not alter the archived primary data.

The long-term value of the databases we have examined is severely limited by inattention, in at least some cases, to the importance and nature of metadata. For example, the idea that a database can contain the initials of the individual collecting the data and that future investigators should contact that individual in order to determine the

characteristics (treatment history) of specific batches of fish is shortsighted. In a few years most such people would have moved on to other responsibilities, retired, disappeared or even died, and the useability of archived data is severely compromised.

Archivists should practice manipulating their data to ensure easy access by users. This would not constitute “analysis”, but degrees of data assembly. Retrieval, collation, and visualization are all aspects of maintaining a useable collection of primary data. Restrictions on such data manipulation (i.e., of Streamnet, PTAGIS (PIT tag database), and RMIS (CWT database)) reduce the value of the data archive to legitimate users.

For each "class" of primary, measured data there needs to be a single archive responsible for those data, in order to guarantee data integrity and to preclude ambiguities. But there can still be many archives, each for a different class of data, as long as each respective data class is officially housed in only one archive. There are in fact advantages to having multiple archive sites, each specializing in one data class. The specialization allows for the staff of the archive site to have better understanding of the nature of their data, of the activities that generate their data, and the kinds of searches and sorts that users would be interested in performing. This should encourage better QA/QC and better communication with the data providers. Note in particular that communication between the data archiving center and the data providers is essential to ensure completeness of the compilation. Finally, specialization among the archive sites will allow for more transparent bookkeeping that could provide an opportunity to control the costs of the archiving function.

It would make good management sense to separate the funding of data input from the funding of the archive site. The data input activity should be budgeted into the respective data generating projects. Delivery of the data, in an appropriate form, and on a specified timetable, should be a contracting requirement for each project that generates data. Needless to say, such a contracting requirement, and separation of the funding of data input and data archiving, will greatly facilitate the establishment of new data archiving sites that are needed to accomplish the data archiving function for the important data classes that, at the moment, are without appropriate archives (e.g. radio telemetry data).

There will be a minimal level of standardization required in the query system of each data archiving site, in order to ensure that the data in each can be accessed by a common query system. Such standards are now common knowledge within the information technology community, and should not constitute an undue burden to the archiving sites. The data archives should be separate (at least conceptually) from data access sites (often called second-tier databases and are discussed further below).

The data archiving system as a whole should be managed through the project funding process, treating each data archiving site as a project identified with serving a specified needed function.



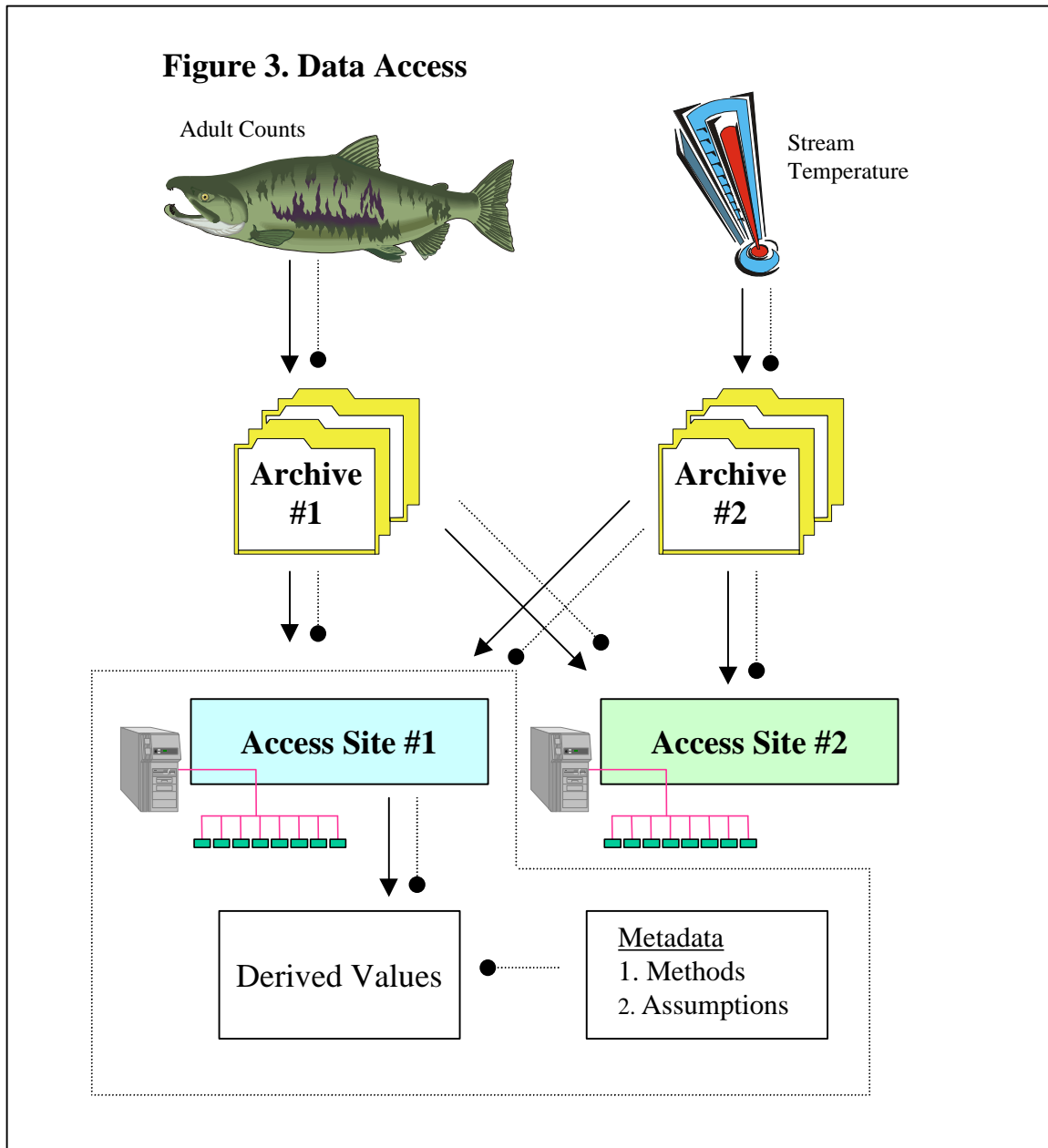


Figure 3. Data Access. Access to data is facilitated by sites that maintain systematic approaches to obtaining data needed for analyses from the data archives, and maintain (or can easily recreate) derived information. For example, two data archives (adult fish counts, temperature) are accessible to several access sites, two of which are shown. Each access site retrieves both primary data and metadata (shown by coded arrows, as in figures 1 and 2) and creates derived values (detailed for Access Site #1). The derived values also have associated metadata showing how these values were derived.

## DATA ACCESS

Data access sites (sometimes called second-tier databases) provide reasonably convenient and effective means to search, sort, and collate data -- possibly from multiple data archives -- for purposes of analysis (Figure 3). They retrieve (copy) data from the data archives (e.g., PTAGIS, Streamnet, the Fish Passage Center, Corps of Engineers, EPA, etc.), but this leaves the archives intact and unchanged. Analytical methods and results of analyses (derived quantities) may then be stored in an access site's database. For example, an access site could produce graphs of accumulated data or calculations of project specific mortality rates derived from PIT-tag data. StreamNet provides a limited menu of such functions, while DART and the Fish Passage Center provide different menus and custom analyses. Although not automated in the sense described here, data analysts in the region have provided derived data from multiple sources to reconstruct (estimate with the aid of various assumptions and models) run sizes for various index stocks. It is surprising that run reconstruction (estimation) does not seem to be a regular part of any of the data access sites reviewed, though some results of run reconstruction exercises are archived (without adequate documentation).

Just as specialization of data archiving centers could provide a benefit in encouraging a tighter bond between data generators and the data archive sites, specialization of data access sites could provide a benefit in encouraging a tighter bond between data users and the data access sites.

The measure of adequacy of a data access system is a moving target. The perceived adequacy of a data access system is evaluated relative to the needs of data users. When new kinds of analyses are undertaken, this can create new data access needs that may not be well served by existing query systems. Data access systems (StreamNet, DART, the Fish Passage Center and private contractors) should be encouraged to meet the needs of all data users in the region including private groups.

Base funding of multiple data access centers might be viewed as duplication of effort, but in the opinion of the ISRP, some duplication of effort is desirable. Competition among the data access centers to provide service to data users provides an added benefit to the region. Furthermore if centers are allowed to charge for customized services, then unique needs of the region might be met in a more cost and/or time efficient manner. For example, project managers could be allowed to subcontract with the data access center that is most cost and time efficient in meeting their needs. This is not meant to interfere with base services and public access to data. An added potential for efficiency would arise when a customized service, whose development and implementation is funded by a particular user constituency, gets incorporated into the menu of regular features at that data center, so that future use does not incur special costs.

The functionality of a data access center seems to be enhanced if the center is tasked with providing custom analyses to data users including private organizations, e.g. DART and the Fish Passage Center. The ISRP believes that functionality of Streamnet could be

improved if they were assigned a larger role as a data access center and asked to provide custom analyses for users including private groups. For example, all of the data access centers should be able to access, assemble, collate, and summarize data from the U.S. Forest Service archives (see the Interior Columbia River Basin Ecosystem Management Plan (ICBEMP)).

The ease of establishing links on the web also extends to links between data access sites and multiple data archives. For this reason there is no intrinsic motivation for the Council to fund duplication of actual function among access sites, since each can readily link to another for a query function that has already been developed. If DART or the Fish Passage Center has been funded to provide a function based on data from the archive site, PTAGIS, (e.g., the mark-recapture histories of PIT-tagged juveniles during downstream migration), then it makes good sense to meet that data access need through a link to their site. The only reason to invest in implementation of a particular new query function within any data access site is to address a need that is not yet met conveniently in the system.

This incremental evaluation of the need to implement a new query would seem to lend itself to an RFP process whereby identified data access needs are described in detail from the perspective of the data user group, and put out to bid. For example, the run reconstruction function mentioned above might be requested in an RFP and be made part of DART, the Fish Passage Center, or Streamnet.

Another area of need that might be put out for bid is to document methods used to obtain derived data. The data user must be able to understand the source and derivations, i.e., the metadata. For example, there are many different graphs and tables of smolt-to-adult return rates (SARs) being circulated in the region with no adequate documentation on methods and models used to derive them (Peter Karieva, NMFS, personal communication). SARs are among the most common “data” quoted in public discourse and it is unfortunate that methods to obtain these derived quantities are not better documented.

The bids may come from existing data access sites or from proposed new ones. This should prove more efficient than attempting to create one highly centralized one-size-fits-all data access site, where the connection between features and actual user needs could get lost, and where the accounting for the cost of serving each user group would be harder to track.

Data access centers may generate derived quantities. The process of generating derived quantities involves assumptions and models. For this reason it is imperative that the associated metadata be maintained with the derived values. In addition, the derived quantities are inherently subject to additional uncertainty owing to the use of primary data that are subject to measurement and sampling variance. It is desirable to quantify the total uncertainty in order to have some sense of what weight the derived quantity should be given if it is used for decision purposes. For this reason, it is good practice to report

the quantification of uncertainty whenever reporting the derived quantity, and to store these numbers in association with the derived values and the metadata.

One standard for adequacy of the metadata (documentation) for derived quantities is whether they provide enough information to calculate a credible quantification of the uncertainty. A second standard for adequacy of metadata for a derived quantity is whether it provides enough information to recalculate the derived quantity from the underlying primary data. If these criteria are not met, the metadata are not adequate.

It may be convenient, but it is not absolutely necessary to archive derived quantities, because with adequate documentation they can always be reconstituted from the archived primary data and metadata. Indeed, it is best that any archiving of derived quantities be flagged, so that users do not mistake them for primary data. Furthermore, it is generally worthwhile to revisit the calculation of derived quantities, at intervals, since assumptions may be revised with experience, and modeling and analytical methods may evolve.

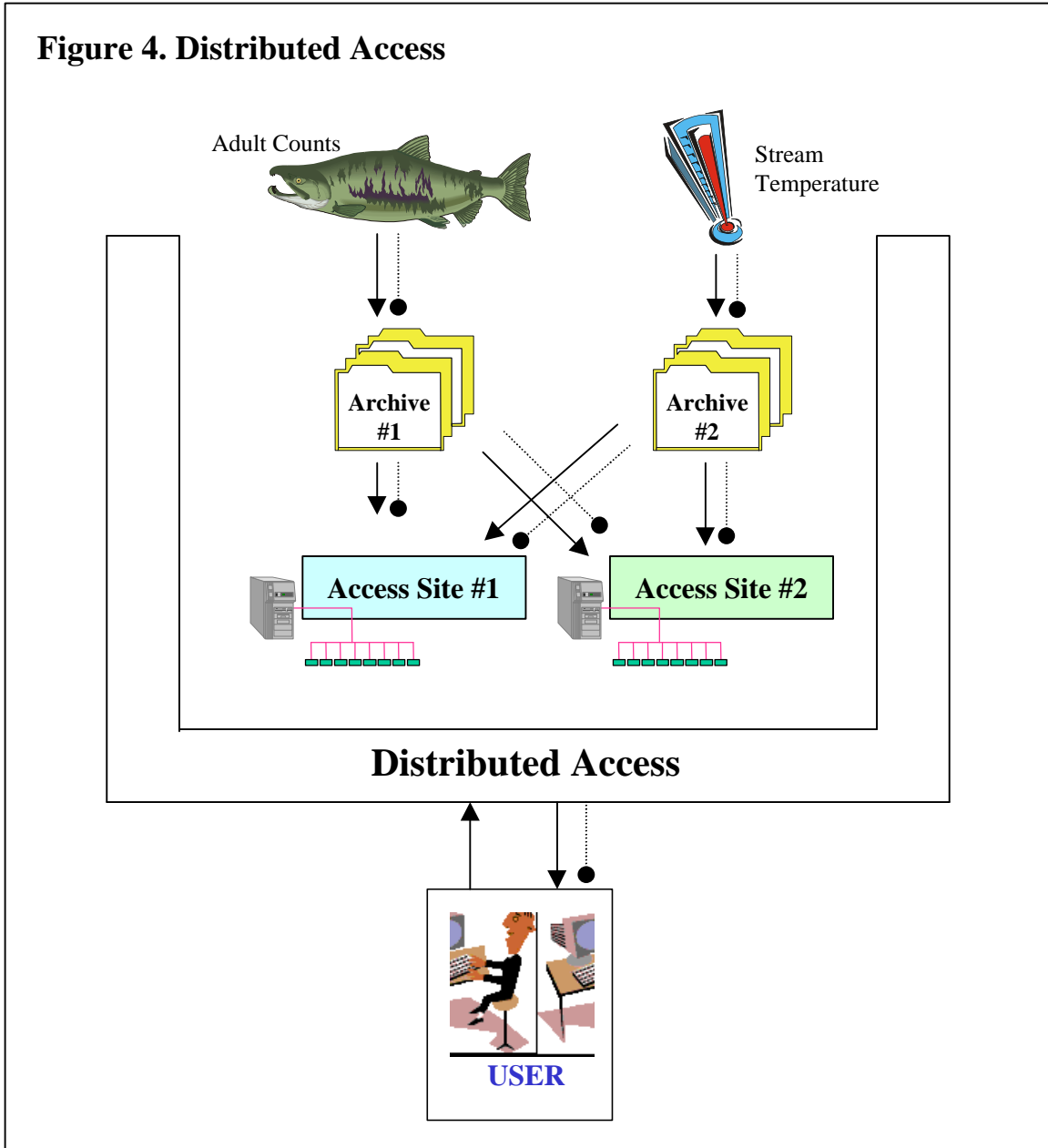


Figure 4. Distributed Access. As a further expansion of Figure 3, the sets of data archives and access sites can be fully encompassed and linked via the Internet as a “distributed access system” shown by the U-shaped border. A user can enter the system through a special distributed access portal or through any of the access sites. A distributed access portal usually maintains special software that facilitates navigation of the desired archives and access sites.

## **Distributed Access**

Scientific data management is now evolving to distributed access systems using the Internet (Figure 4). This is known as a “distributed environment.” All of the data, documentation, and metadata (indexing information) can reside with the individual data providers’ servers. A distributed access system uses the Internet to form a “virtual system” interconnecting those servers and its central system. The metadata are saved in sections, known as “XML data islands,” inside documentation files that are in HTML format in the system’s server. The metadata are automatically extracted, or “harvested,” to the central system periodically (as often as daily for rapidly updated files). The actual data are controlled by the providers. A user needs only query the central system in order to be routed automatically to the appropriate peripheral server or servers.

There are several advantages to a distributed system. For the user, the main advantage is one-stop, central access. With a distributed system, the central system does not need vast resources to maintain a copy of all data. Hardware costs are limited, and cost savings are passed on to users. The system allows the primary data archives (such as a fish-tag database or individual researcher) to control which data sets, metadata and documentation files are shared with the system. The archivist (or researcher) creates the data sets, provides quality control, and allows data sharing by providing the system with a list of accessible URLs known as locator files. The files to be shared can be changed (added, deleted) at any time at the will of the primary archivist. Depending on the policy of the project, researchers may choose to protect their data by placing them in an area with password restrictions.

The distributed system’s low level of effort and minimal disruption to data providers are two of its most attractive features. Many projects are already required to publish their results on the Internet. Running a Web server or using a Web hosting service is a task with which most projects are already familiar. The distributed access system adds few new requirements to a Web system that is already in place. The system simply requires that the metadata be placed in HTML or XML files (both are text based) using agreed-upon metadata standards and that the files be available on the World Wide Web.

The system staff or staff of a data access site provides tools to help users find and create the files they want for their job at hand, i.e., find, sort, and collate data for purposes of analysis. This can involve more than one primary data archive and more than one specialized data access site. Metadata editors create composite files for a user with the format and codes automatically included. The user only needs to specify which data sets are wanted and the methods to be used.

There are several distributed access models that may be useful for integrating data sets in the Pacific Northwest. DART has evolved into a distributed access system at the University of Washington’s Columbia Basin Research office (<http://www.cbr.washington.edu>). The Fish Passage Center is performing the role of a data access center using the region’s database resources in a distributed manner

(<http://www.fpc.org>). The Council and CBFWA are establishing a site (the Columbia River Ecosystem Research, Monitoring and Evaluation data directory (the “yellow pages” site)) to aid in location of past and current projects (including data and reports) conducted in the region. A federal interagency system (called “Mercury”) was developed at Oak Ridge National Laboratory and is used by a number of federal and international agencies (<http://www.mercury.ornl.gov>). Because of the low cost of a distributed access system, there is no need to have only one operational system for a large amount of data. Multiple systems allow for competition for providing the most advantageous user services, in both technology and staff assistance.

A distributed access system is most easily implemented given standardized methods and metadata describing the information contained in each data set. The Columbia River Basin does not yet have standardized methods and this makes access to data more complex. Further, data archive sites do not currently provide all necessary metadata for their primary data. These are perhaps the biggest obstacles to implementing fully functional distributed access systems in the near term.

## Appendix 1. Assignment from the Northwest Power Planning Council

**FRANK L.  
CASSIDY  
JR.  
"Larry"  
CHAIRMAN  
Washington**

**NORTHWEST POWER PLANNING COUNCIL  
851 S.W. SIXTH AVENUE, SUITE 1100  
PORTLAND, OREGON 97204-1348**

**ERIC J.  
BLOCH  
VICE  
CHAIRMAN  
Oregon**

**Tom Karier  
Washington  
Todd  
Maddock  
Idaho**

**Fax:  
503-820-2370**

**Phone:  
503-222-5161  
1-800-452-5161**

**Internet:  
www.nwppc.org**

**John Brogoitti  
Oregon  
John Etchart  
Montana**

**Mike Field  
Idaho**

**Stan Grace  
Montana**

March 1, 2000

### MEMORANDUM

**TO: ISRP, Rick Williams and Lyman McDonald**

**FROM: Chip McConnaha and Erik Merrill**

**SUBJECT: ISRP Database Review**

The ISRP's past two June 15 reports have included recommendations for an independent review of data management efforts that are funded through the Columbia River Basin Fish and Wildlife Program. Specifically, the ISRP recommended that this review take place before funding is continued beyond FY2000. Over the past few months, Council staff and the ISRP have agreed that the ISRP should undertake this review. The Council takes particular interest in this subject and has designated Washington Council member Dr. Tom Karier to coordinate a comprehensive review of data management for the Columbia River Basin. The Council expects the ISRP report to be an important element of this review. This memo is intended to assist the ISRP as it scopes the review to focus on issues that will be most helpful for Council decisions.

From the ISRP review, the Council primarily is interested in:

1. An expansion of the ISRP's observations on database management that were described in the ISRP's past reports. For those reports, the ISRP reviewed proposals for PTAGIS, the Fish Passage Center, the UW Data Center (DART), BPA's Second Tier Database, and StreamNet. The ISRP reported that it could not find compelling reasons for funding multiple databases of smolt monitoring results. The ISRP found much



potential for duplication of services, little justification for how an individual data system is distinct from the others in its scope or in the services it provides, a lack of coordination, and the potential for different versions of the “same” data, which could lead to confusion. In addition, the ISRP observed that to the extent that the databases are not duplicative, data users are not well served. Do these projects supply the region with information needed to plan and manage the Columbia River ecosystem? What are their strengths and weaknesses? How do they address the needs of the diverse audience including researchers, managers and decision-makers?

2. General recommendations on how to provide the region with a geographically referenced information system that facilitates timely application of information collected under the Council’s program and includes related information from other programs. Such a system should focus on ecological features of the basin and its habitats but include information on biological performance of fish and wildlife. What type of data structure is needed to ensure that the necessary data is accessible for analyses to inform decision making in the basin?

In addition as PATH transitions out, the Council would welcome advice from the ISRP on how to ensure the continuation and improvement of structure analytical functions such as collaborative analysis, maintenance of derived data, data mining and validation.

The ISRP report would be of the most value to the Council if completed by May 15, 2000.

## **Appendix 2. Brief overview of projects reviewed, summary of ISRP briefings, and documents presented to the ISRP for review**

### **1. Data Analysis in Real Time (DART); [www.cgs.washington.edu/dart](http://www.cgs.washington.edu/dart) (BPA Project Number 9601900) Second Tier Database Support for Ecosystem Focus**

On February 8, 2000, Jim Anderson and John Skalski briefed the ISRP Subcommittee on the “Second Tier” database DART (Data Access in Real Time) maintained by BPA at the University of Washington, Columbia Basin Research office, and provided a demonstration in the Columbia Basin Research office.

Objectives: Provide a single-point source for Fish and Wildlife Program participants needing integrated current and historic information from more than one primary data provider.

Services Provided: Provides access to current and historic information from other sources, such as StreamNet, the Fish Passage Center, and others. A Report Generator allows the user to select one or more routinely prepared documents, graphs, etc. for viewing and perhaps printing, as a single package. The separate items may reside anywhere on the Internet.

Documents received:

- hard copy of slides
- Columbia River Basin Data & Web Products – Survey

### **2. The Fish Passage Center (FPC) (9403300); [www.fpc.org](http://www.fpc.org)**

On March 21, 2000, Michelle DeHart (Director), Henry Franzoni, and Penelope Sanders briefed the Subcommittee on the three components of the Fish Passage Center: the Fish Passage Center Project, the Smolt Monitoring Project, and the Comparative Survival Study Project, all of which have data management as an essential element.

Objectives: The Fish Passage Center provides the fishery agencies and tribes with technical expertise on hydrosystem operations, smolt monitoring data used for fish passage management decisions, and management of a database with fish passage information. The Fish Passage Center designs and oversees the smolt monitoring program which includes counts of juvenile salmon at bypass samplers and checking for gas bubble trauma, maintains daily records of hatchery releases, hydrologic information, power operations, reservoir operations, adult passage counts, water temperature and dissolved gas levels. Historic information is maintained in databases. The databases provide data recovery for numerous studies basinwide and are available to all public and private entities. An Internet web page is maintained.

Services Provided: The Fish Passage Center is supervised by a Board of Directors from CBFWA. As requested by the Board of Directors and the Anadromous Fish Managers (comprised of representatives of the state and federal fishery agencies and treaty tribes of

the Columbia Basin), the Fish Passage Center provides specific analysis of alternatives for fish passage, such as may be useful in decisions on flow augmentation, spill, adult passage, and the like. The Fish Passage Center staff provides analysis and reports to state water quality agencies as requested by the agencies and tribes. The Fish Passage Center is a key component of the Smolt Monitoring Program (SMP) and the Comparative Survival Study (CSS). The Fish Passage Center designs and oversees the SMP, and is responsible for management of the CSS.

Documents received:

- Fish Passage Center Data Catalog (3/21/00)

### **3. Smolt Monitoring by Federal and Non-Federal Agencies (SMP) (8712700)**

Objectives: The SMP provides counts of outmigrating juvenile salmon in bypass samplers at key projects on the Snake and Columbia rivers. PIT tagged fish are recorded as they pass the projects. Recording of the PIT tag recovery data is accomplished through PTAGIS. Dissolved gas monitoring takes place according to NMFS Dissolved Gas Monitoring Plan. Rate of descaling is recorded. Determination of the timing of the outmigration is made from the counts. Data are transmitted daily to the Fish Passage Center where decisions on operations of the hydropower system are made on the basis of the information gained. These include implementation of measures in the Biological Opinion relative to flow and spill management, especially during periods when spill is provided to improve smolt survival. PIT tag recovery information is used to estimate rate of migration and survival through river reaches

Services Provided: Monitoring sites are Lower Granite, Little Goose, Lower Monumental, McNary, John Day, Bonneville, and Rock Island dams, and the Lower Grande Ronde River trap. While the project is overseen by the Fish Passage Center, the BPA contracts are with Washington Department of Fish and Wildlife, Idaho Department of Fish and Game, Oregon Department of Fish DNA Wildlife, Nez Perce Tribe, and Chelan County P.U.D., whose personnel conduct the on-site monitoring.

### **4. Comparative Survival Study (CSS) (8712702)**

Objectives: The Comparative Survival Study analyzes PIT tag recovery data to compare survival estimates for transported fish of known origin with survival rates of downriver stocks, wild and hatchery transported fish and fish handled and not handled at dams. Comparisons are based upon smolt-to-adult survival indices for spring and summer stream type chinook originating above Lower Granite Dam to evaluate mitigation measures and actions, such as flow augmentation, spill, and transportation that are undertaken to benefit salmon stock listed under ESA.

Services Provided: The study is designed to provide information the PATH process recommended as being needed to address the question “can transportation of fish to below Bonneville Dam compensate for the effect of the hydro system on juvenile survival rates of the Snake River spring and summer chinook salmon during their downstream migration?” Groups of spring chinook salmon from Lookingglass, Imnaha,

McCall, Rapid River, Dworshak, Round Butte, Warm Springs, and Carson hatcheries are PIT tagged as part of the regional Smolt Monitoring Program. An interagency committee, the Comparative Survival Study Oversight Committee analyzes the recovery of tag data in returning adults, resulting in smolt-to-adult survival estimates, comparison of wild and hatchery chinook SARs, evaluation of the transportation program, and comparison of upriver and downriver SARs. The ISAB reviewed and approved the study proposal and design. The ISAB in 1997 recommended consolidating the study under one project number, which is being explored.

Data on tag recoveries are obtained from PTAGIS for analysis in the Comparative Survival Study. Results of analyses are available through the Fish Passage Center and through BPA.

### **Pacific States Marine Fisheries Commission; [www.psmfc.org](http://www.psmfc.org)**

On March 21, 2000, the ISRP received a briefing on PSMFC and its history of housing several databases for the region by Randy Fisher, Director. Briefings for StreamNet, PTAGIS, and the Coded-Wire Tag database, each administered by PSMFC, were held at PSMFC offices (see below). Attendees (other than presenters listed below) included John Picinni (BPA COTR for StreamNet), Pat Poe (BPA COTR), Doug Taki (PIT Tag Steering Committee), Phil Roger (CRITFC), Cedric Cooney, Dick O'Connor, Karen McGill, and Stan Allen.

### **5. PIT Tag Information System (PTAGIS) (900800);**

[www.psmfc.org/pittag](http://www.psmfc.org/pittag)

On March 21, 2000, Carter Stein, project leader, and Dave Marvin, biologist, described the database for all PIT tag information.

Objective: Develop, operate, maintain and enhance a long-term Columbia River Basin database on PIT tag information and provide operations and maintenance support for the collection of PIT tag information at PIT tag interrogation sites.

Services Provided: PTAGIS provides database systems management and operations for the collection and distribution of PIT tag data to all interested parties. PTAGIS provides for the operation and maintenance of transponder interrogation systems located at juvenile bypass facilities at hydroelectric dams on the Columbia and Snake rivers. The program supports the Pit Tag Operations Center (POC) which maintains software and documentation and provides user support and training for PTAGIS related tools. A steering committee, the Columbia River Basin Pit Tag Steering Committee, works to establish data standards, system development priorities, and operational priorities with respect to the operation and maintenance of the interrogation systems. PTAGIS performs no analysis of data other than to determine coil and monitor efficiencies for systems operation.

The PTAGIS project distributes three user documents to facilitate access to the system: 1) the PIT Tag Source Data Input Specification; 2) The PIT Tag Tagging Procedures Document; and 3) The PTAGIS Database Users Manual.

The database is used by many PIT tagging operations. A web site is maintained for public or casual users, while researchers are provided a login code giving them ready access to the database. Other data systems may access the PTAGIS database and utilize the information for their analyses and decision making. These include the Fish Passage Center and StreamNet. StreamNet provides links to PTAGIS but does not itself include any PIT tag data.

Justification: PSFMC was asked to undertake this task in 1990 when NMFS announced its intention to withdraw from management of the PIT tag data system. PSMFC was asked for several reasons, previous experience in management of data of this kind, and its independent position with no vested interest in the interpretation of data generated.

Documents received:

- PTAGIS (hard copy of viewgraphs and notes for ISRP presentation)
- PTAGIS Response to Questions from ISRP
- PTAGIS Newsletter, 4(1) March 2000 (sample of what's on the Internet)
- Charter, Passive Integrated Transponder (PIT) Tag Steering Committee
- Bauman, L., J. Mead, and C. Stein. 1998. PIT Tag Information System User Manual. Version 4. Pacific States Marine Fisheries Commission, Portland, Oregon.
- PIT Tag Steering Committee. 1999. PIT Tag Marking Procedures Manual. PITTag2. Software User Manual. PIT Tag Operations Center, Columbia Basin Fish and Wildlife Authority, Portland, Oregon. 1999.
- Stein, C., D. Marvin, J. Tenney, and L. Bauman. 1999. 1999 PIT Tag Specification Document, PIT Tag Information System, Columbia River Basin. Pacific States Marine Fisheries Commission, Portland, Oregon.

## **6. Coded Wire Tag Recovery (CWT), Regional Mark Information System (RMIS) (8201300); [www.rmis.org](http://www.rmis.org); [www.psmfc.org](http://www.psmfc.org)**

On March 21, 2000, Ken Johnson, project leader, and Jim Longwell, computer specialist, discussed the Regional Mark Processing Center for coded-wire tag information that maintains the Regional Mark Information System for coded-wire tag data.

Objectives: To sample statistically valid numbers of chinook and coho in the Columbia River and Oregon and Washington coastal communities and recreational fisheries and to estimate the escapement. This project falls within the Coded Wire Tag umbrella project, and is responsible for sampling to recover coded wire tags from the fisheries and escapement areas, and management of the resulting data.

Services Provided: The states of Oregon and Washington jointly carry out a sampling effort (minimum 20% sampling rate) to collect coded wire tags from the fisheries and escapement areas. Tag recovery labs are located at Clackamas, OR and Olympia, WA

where the tags are sent to be decoded. The resulting data, including the tag code, and recovery and catch/sample rate, are forwarded to PSMFC's Regional Mark Processing Center where the inputs are validated and made available to users via the on-line "Regional Mark Information System". The information has numerous applications and is employed by many users.

Coded wire tags are used coastwide for various purposes from California to Alaska. All of the agencies participating in the PSFMC contribute to the budget of the Coded Wire Tag Recovery Project. BPA funds what is considered to be a fair share of the budget, based upon the relative numbers of tags applied for studies within the Columbia Basin.

Documents received:

- Regional Overview of the Coastwide CWT System. Prepared for the ISRP March 21, 2000.
- CWT Data File Definition, Specification, & Validation. Version 3.2. Regional Mark Processing Center, Pacific States Marine Fisheries Commission, Portland, Oregon. June 12, 1998.

**7. StreamNet, The Northwest Aquatic Resource Information Network (8810804); [www.streamnet.org](http://www.streamnet.org)**

On March 21, 2000, Bruce Schmidt, project leader, briefed us on the database history and current status. Lenora Oftedahl, librarian, briefed us on the StreamNet Library housed at CRITFC offices. StreamNet is administered by the Pacific States Marine Fisheries Commission, with many agency and tribal cooperators. It was formerly known as the Coordinated Information System (CIS) and Northwest Environmental Data Base.

Objective: To coordinate systemwide data collection systems, promote exchange and dissemination of information having a bearing on the Fish and Wildlife Program of the Northwest Power Planning Council, in a standardized electronic format throughout the Columbia Basin.

Services Provided: The StreamNet Distributed System is a PC based database application containing fully referenced data and an interface to query, report or export the data. The services provided by StreamNet are designed not to overlap those of Fish Passage Center, CWT or PTAGIS. StreamNet maintains a library and reference system to ensure that data are referenced and that important reports are available for use in monitoring and evaluation of Columbia River fish stocks. The 1994 Fish and Wildlife Program calls upon CIS (now StreamNet) to provide an annual program monitoring report. StreamNet provides an annual report on the status of fish runs by bringing together many diverse data types, including data on environmental conditions that may affect that status. This synopsis is intended to complement other more detailed reports to which readers are referred for comprehensive treatment of specific subjects. Available information is identified with no attempt to evaluate its implications.

Documents received:

- Answers to ISRP Questions
- StreamNet Data Summary (table)
- StreamNet Province/Subbasin Data Inventory (draft March 2000)
- Listing of .gov, .us, and .mil domains which accessed StreamNet during February 2000
- StreamNet Library Access Guide

## **8. Bonneville Power Administration's Columbia Basin Fish and Wildlife Information Portal**

On March 22, 2000, Dave Askren, BPA Fish and Wildlife, provided the subcommittee with BPA's perspective on regional management of fish and wildlife information, including ideas for a Columbia Basin Fish and Wildlife Information Portal, existing and future policies and contract requirements, and document management practices.

Documents received:

- Regional Management of Fish and Wildlife Information – BPA Perspective (copies of slides and outline)
- Bonneville Power Administration Policies and Implementation Affecting Information Management

## **Northwest Power Planning Council**

On March 22, 2000, the ISRP met with Tom Karier, Washington Council member. He described his interest in database management and attempts to obtain more information from contractors when they submit annual requests for funds.

## **9. Columbia River Ecosystem Research, Monitoring and Evaluation; [www.cbfwf.org/find](http://www.cbfwf.org/find)**

On March 22, 2000, Gustavo Bisbal (NPPC) and Eric Schrepel (CBFWA) briefed the subcommittee on the progress of a recently initiated distributed information system for regional fish and wildlife research and monitoring projects. This new tool is intended to link existing databases through “metadata” on each data management center. The analog is a “yellow pages” directory of available data. It would make available, in a single source, access to all databases and other sources of information relevant to implementation and evaluation of the Council's Fish and Wildlife Program.

Services to be Provided: While this system is currently available to all users through the Internet, it continues to be developed. The system is not for archiving primary data, but to direct the user to the appropriate source for the data of interest. Searches should be possible through different fields, e.g. by agency source, investigator's name, location within the basin, species, stock and other subject categories. Initially, the proposal is to incorporate Fish and Wildlife Program projects currently underway. Eventually, the plan would be to include historic data.

Documents received:

- Answers to ISRP questions

## **10. Northwest Power Planning Council Framework and Ecosystem Diagnosis and Treatment (EDT)**

On March 22, 2000, Drew Parkin discussed the Framework and EDT processes in relation to the database developed for the EDT model. While EDT is primarily a modeling approach aimed at estimating potential salmon production of subregions in the Columbia River Basin, the approach has required assembly of a database that includes a wide variety of primarily derived environmental quantities. The “EDT” database will eventually be made available for further uses with access through the Internet.