Introduction

Xanthippe Augerot

As of 2003, the Order *Salmoniformes* (i.e., salmon, trouts, chars) contained 217 valid taxa (Integrated Taxonomic Information System, <www.cbif.gc.ca>); an additional species was described in 2005 (Safronov and Zvezdov 2005). While this book has applications to the full array of salmonid taxa, it is based largely on the science developed on Pacific salmon.

Pacific salmon inhabit streams stretching from California to Japan, encompassing some 31 million square kilometers and spanning hundreds of jurisdictions. They begin their complex life history in freshwater, migrate and mature at sea, and return to their natal streams, where they spawn, die, and deliver vital ocean-derived nutrients to freshwater, riparian, and terrestrial ecosystems. Few fish species are as important to North Pacific ecosystems, native peoples, and coastal economies as Pacific salmon. Their role in nature and value to humankind inspires intensive and frequent study. Whether surveying salmon on the Kamchatka peninsula in Russia or in the Columbia River basin in the United States, scientists need a set of standard monitoring protocols that minimizes methodological errors, maximizes the validity and consistency of data, and allows them to make reliable comparisons and reasonable conclusions across projects and river basins and over time. This book's objective is to describe a standard set of monitoring protocols and best practices that decision makers and funding organizations can adopt and practitioners can use to design study and sampling techniques, conduct field activities, and manage spatial and tabular data.

Background

The Wild Salmon Center (WSC) and the joint WSC–Ecotrust State of the Salmon Program recently concluded its first range-wide assessment of the risk of extinction for Pacific salmon (Augerot 2005). A formidable challenge was trying to piece together fragmented species data across jurisdictional boundaries, understanding that differences in data collection methods and the uneven distribution of data by species and region were largely responsible. Similar difficulties have been encountered in the Pacific Northwest region of the United States. Government decision makers at the Bonneville Power Authority and the Northwest Power and Conservation Council and members of the United States Congress have spent hundreds of thousands of dollars over the years to track salmon status and temporal trends across the Columbia River basin only to realize that individual projects cannot easily be "stitched together" due to inconsistent approaches used to collect and count fish.

To address these fundamental differences in data collection, David Johnson, then at the Washington Department of Fish and Wildlife, proposed bringing together a group of regional experts to determine best practices for a suite of commonly used field techniques and to merge the results into an easy-to-use resource. In 2004 State of the Salmon and the Washington Department of Fish and Wildlife coconvened a workshop of international experts in Welches, Oregon to discuss a monitoring strategy for Pacific Rim salmonids. Workshop participants reviewed 375 documents and took a major step towards producing standardized monitoring protocols that could be used in the field. Following the workshop we whittled down, refined, and standardized the array of field monitoring techniques to 18 core techniques that were then subject to extensive and intensive review by our peers. This handbook is the culmination of these efforts and represents the most comprehensive set of monitoring protocols ever compiled for Pacific salmon and trout.

Meeting the needs of policy makers

Utilizing best monitoring practices is an essential first step to estimating salmonid population status and prioritizing conservation actions. Decision makers and fishery managers find project-level data most useful if field research and monitoring objectives are clearly defined. One of the most frequently overlooked elements of salmon monitoring is selecting appropriate sample design—how we space samples over habitats and over time—to answer relevant questions. Sample designs need to yield the best possible data in a timely and cost-effective manner to be of maximum value to decision makers. While key research and monitoring questions may seem straightforward on paper, in reality the process is often complicated. This is especially true in regions where salmon populations straddle land ownerships and site accessibility becomes an issue. Competing demands for land, water, fish, and wildlife can also create friction among user groups and put researchers in the potentially awkward position of limiting or altering the scope of studies for political reasons. The Columbia River basin in the Pacific Northwest embodies many of these challenges.

The data collection and enumeration protocols in this book serve two distinct monitoring and management goals: to support sustainable salmon harvest and to ensure the viability of threatened and endangered salmon populations. In Alaska, northern British Columbia, and much of the Russian Far East, the principal salmon monitoring objective is stock forecasting, to allocate fishing effort in a manner that allows for sustainable salmon productivity in perpetuity. In shorthand, we can refer to this monitoring and management approach as monitoring to optimize biomass (Hyatt 1996). As one progresses farther south into British Columbia and the U.S. Pacific Northwest, where river basins have been more heavily altered by people, endangered species mandates such as Canada's Species at Risk Act (SARA, 2003) and the U.S. Endangered Species Act (ESA, 1973) drive another monitoring mandate: to provide status assessments, trend assessments, and assessments of salmon population viability across conservation units.[1] For the sake of convenience we will refer to this mandate as managing for biodiversity conservation. Last, the research community is amassing evidence to support what many long suspected, that the annual return of the salmon to our North Pacific river systems is a vital element of the nutrient supply system, feeding riparian vegetation, insects, fish, and birds and supporting ecosystem health (Gende et al. 2002). Canada's new Policy for Conservation of Wild Pacific Salmon explicitly acknowledges maintenance of ecosystem integrity as one of its three objectives.

As salmon are listed as threatened or endangered after experiencing precipitous declines in distribution and abundance, additional legal protections are triggered under endangered species laws. However, we find time and again that we do not have the basic building blocks of conservation knowledge to support salmon recovery efforts. We lack information about the distribution of the dozens of discrete biological populations across the landscape. Traditionally, the larger, mixed population fishery "stock" units have been used to gather information

[1] The European Union Directive on the Conservation of Natural Habitats and of Wild Fauna and Flora provides a similar biodiversity mandate (Cowx and Fraser 2003).

and infer status, but the information these units offer is too coarse. As a result, we cannot characterize populations in terms of demographic performance (e.g., recruit-per-spawner, fecundity, survival rates, size-at-age, sex ratios) and life history diversity. We rarely know which populations in a given conservation unit, or evolutionarily significant unit (ESU) for salmon in the United States), are the most productive and represent the strongholds for conservation and restoration. To improve our understanding of population productivity and its bottlenecks in freshwater or at sea, we particularly need low-cost, reliable means to estimate smolt out-migration. Such knowledge about relative population productivity will be vital to designing effective networks to protect wild salmon productivity in salmon conservation reserves.

The Columbia Basin conundrum

In the Columbia River basin, where the initial idea for this book arose, policy makers operate under both the biomass and biodiversity mandates, which are not always in alignment. Agencies setting policy face the need to balance fish and wildlife conservation with a major hydropower system, tribal treaty fishing rights, commercial fishing interests, and aquatic habitats affected by a growing human population. For instance, salmon hatcheries reflect the complexities of the biomass-versus-biodiversity debate. Hatcheries, mandated by the federal government as a mitigation strategy for habitat loss, account for much of the salmon caught by commercial and tribal fishermen; however, hatcheries conflict with biodiversity objectives because they introduce large numbers of artificially cultivated fish to the system, creating the perception that wild populations are healthy and masking declines in natural population—the backbone of salmon diversity and resilience.

In addition, the Columbia River basin is a web of jurisdictions and competing sovereign interests and responsibilities. Two countries, multiple tribal nations, dozens of government agencies, and thousands of land ownerships exist within the Columbia River basin, one of the most complex salmon monitoring and management regions in the North Pacific.

A typical menu of field research and monitoring questions for the Columbia River basin may include the following, among other themes:

- Which populations are recovering sufficiently to delist ESUs listed as endangered or threatened under the U.S. Endangered Species Act?
- Which populations are robust enough to tolerate recreational and commercial fishing?
- Are hatchery releases competing with listed wild salmon in the river, in the estuary, and at sea?
- What is the magnitude of hatchery postrelease mortality or sublethal stress associated with passage in-river and trucking or barging smolts?
- What is the incremental benefit in fish survival of retrofitting dams with smolt passage weirs?
- Can we describe the cumulative effect of competition and predation on wild salmon by native predators? By invasive fish species?
- What areas within the basin would be highest priority for habitat restoration, given the presence of highly productive "core" populations?

- Do new mining ventures present a significant threat to salmon abundance, distribution, productivity, and diversity?

To answer any of these questions requires careful consideration of the field research approaches, desired level of certainty for answers, frequency of information needed, and scale of information needed (e.g., ESU, county, whole-basin), among other factors. Research questions evolve into research objectives and methods, the cornerstones of well-defined study plans. When field data and initial study plans are poorly documented or described (e.g., little or no metadata), the resulting analyses are scientifically less credible, leading to questions about how they should be interpreted and used by decision makers. As a community, we should strive to document study objectives, sample design rationale, and all limitations with respect to valid inferences from the data we collect.

We cannot guarantee that management decisions or policies will be crafted using the best data, but we can strive to ensure that the best data are available to decision makers for salmon conservation.

Improving the efficiency of monitoring: A common research template

Last, we must recognize that as resource conditions change and management priorities evolve, we will need to ask a different set of questions and rely on new data gathering and management tools to inform salmon conservation objectives. The salmon research community could collect data, generate new information, and expand its knowledge base more economically if it used the same methods, metrics, and information system for sharing basic but essential data gathered at different sites over varying lengths of time. One approach is a master sample framework, a monitoring template designed for and applied to river basins, such as the Columbia River basin, where the distribution of fish populations is documented. Developed by survey statisticians, a master sample frame is designed to provide a pool of sample locales across an area, ensuring statistical reliability at multiple spatial scales. If a master sample framework were established for Columbia River basin fish monitoring—and maintained and updated by a statistical committee representing the sampling interests of county, state, tribal, and federal biologists and housed in a central data system—it would provide cost efficiencies in survey design and facilitate the integration of data across individual studies (Turner 2003). Larsen et al. (2003) recommend a master sample consisting of a dense, spatially balanced suite of sites on stream networks that could be tailored to meet multiple agency design needs for fish monitoring in Oregon. The State of the Salmon program is compiling an accessible database by cataloging, standardizing, and integrating biological data from thousands of study sites to lay the groundwork for the selection of a master sample at the scale of the North Pacific. This effort will improve our ability to monitor trends in salmon abundance, diversity, population structure, and productivity over the range of Pacific salmon.

We recognize that the goal of creating a useful template for documenting salmon monitoring activities across boundaries and providing a research reference point for future studies is extremely ambitious; however, this is an achievable goal if the salmon research and conservation community can learn to speak the same language. This means that our salmonid management objectives, key research questions, and design of monitoring programs must be consistent, repeatable, and verifiable to yield the best possible information. We believe that this handbook represents a significant step towards developing that common language.

How to use this handbook

The next two chapters address study and sampling design and the implementation of robust data systems. These two chapters provide a framework for the heart of the handbook—the technique-based field protocols. Each adult and juvenile salmon collection and enumeration technique has its own chapter. While each chapter was authored by a different team of field biologists, each technique is presented and described using the same template for usability:

- Background and objectives
- Sampling design
- Field/office methods
- Data handling, analysis, and reporting
- Personnel requirements and training
- Operational requirements (including budget)
- Literature Cited

The background and objectives section provides context for the sampling technique and will help guide researchers to the most appropriate one, depending on the focal species and site conditions. It also describes the evolution of the technique, the most common objectives for deploying the technique, and for which species, age classes, and environments it is best suited.

The sampling design section contains more detailed information about the applicability of the technique to meeting specific research objectives for particular species and environments. This section also describes elements of overall study design and appropriate spacing of samples in space and time. (For example, weirs, smolt traps, and sonar enumeration systems all have specific siting requirements that may determine where and how each is used to accomplish study objectives.)

The field/office methods section provides an overview of field logistics and related office support needed to assist operations at study sites. (In the case of fixed gear such as smolt traps and weirs, field descriptions are very robust and list alternative specifics of gear deployment for a variety of stream conditions.)

The data handling, analysis, and reporting section provides best practices for analyzing field data, for documenting and handling data, and for conducting quality analysis and quality control on resulting data. Some chapters provide detailed discussions about alternative approaches for conducting status and trend analyses.

Sections on personnel requirements and training and other operational requirements further facilitate the comparison across techniques. The roles and responsibilities of personnel and training and safety issues are addressed for every technique. The operational requirements section explains the typical amount of time needed to plan and implement the study and covers field schedule, equipment, facility needs, and budgetary considerations.

Literature Cited

Augerot, X. 2005. Atlas of Pacific salmon. UC Berkeley Press, Berkeley, California.

Gende, S. M., R. T. Edwards, M. F. Willson, and M. S. Wipfli. 2002. Pacific salmon in aquatic and terrestrial ecosystems. Bioscience 52(10):917–928.

Hyatt, K. D. 1996. Stewardship for biomass or biodiversity. Fisheries 21(10):4–5.

Larsen, D., A. R. Olsen, and D. L. Stevens, Jr. 2003. The concept of a master sample: an Oregon case study. Presented at Designs and Models for Aquatic Resource Surveys. Oregon State University, 7–9 September, 2003.

Safronov, S. N., and T. V. Zvezdov. 2005. *Salvelinus vasiljevae sp. nova.* A new species of freshwater chars (*Salmonidae, Salmoniformes*) from northwestern Sakhalin. Journal of Ichthyology 45(9):700-711.

Turner, A. G. 2003. Sampling frames and master samples, draft. Expert Group Meeting to Review the Draft Handbook on Designing of Household Sample Surveys, 3–5 December 2003. United Nations Secretariat Statistics Division, ESA/STAT/AC.93/3.

Evolving towards a Common Global Language for Salmon Conservation

Samantha Chilcote

I recently had the distinct honor to work with Dr. Anatoly Semenchenko, a field biologist at the Russian fisheries agency TINRO Center. Dr. Semenchenko is an esteemed scientist and, for the last several years, has focused much of his efforts on the Samarga River in the southern Russian Far East, largely without institutional support. The Samarga remains relatively untouched, a center of biodiversity in the eastern Sikhote-Alin Mountains and home to one of the world's largest populations of endangered Sakhalin taimen *Hucho perryi*. Dr. Semenchenko was so interested in the Samarga that he gave up his vacation time to join us for an expedition.

As part of my postdoctoral work at the University of Montana and in conjunction with Wild Salmon Center, I was tasked with conducting a rapid assessment of this beautifully complex river and its astounding biotic diversity. It is a 600,000 ha roadless watershed where logging has just begun. Fortunately, the logging company has received Forest Stewardship Council certification and is working with conservation groups to set aside critical habitat and otherwise minimize environmental impacts. Therefore, our scientific research would have direct application on the land and to the fish. I had a lot to learn about this wonderful watershed and had to learn it fast.

Of course, Dr. Semenchenko knows that river better than any fish biologist—but the question nagged at me: what exactly does he know? We studied his earlier rapid watershed assessment; in sum, it is a phenomenal body of information about fish richness, abundance, distribution, and spawning areas. Acknowledging the wealth of information that could be gleaned from a larger, collaborative study, I wanted to tap his experience for the full breadth of what he could tell me.

I asked Dr. Semenchenko about his observations on life history diversity. To me, it was a basic platform: If we do not understand life history diversity, we do not understand what habitats the fish are using. Life history diversity is so often how we frame salmon study. I simply assumed that his observations had the same foundation. But Dr. Semenchenko could not fathom my objectives, and I sensed his annoyance at the possibility that we were planning to impose an entirely new methodology onto his protocols. This went on for about an hour, with a translator as a go-between to try to define what was to me an elemental term. Without clarifying the differences in terminology and the nuances and basic foundations that separate our research, we were stuck. Finally, we determined that what we call life history variations, the Russians call ecological forms.

Of course, my experience in Russia is an extreme example of scientists quite literally using different languages, but we face this problem everywhere at every scale in salmon study—if not in science all around. I am confronting this same lack of clarity here at home right now while working with Wild Salmon Center field biologist John McMillan to refine research and monitoring on the Hoh River in Washington's Olympic Peninsula. We are developing protocols for the Hoh River regarding mapping different floodplain habitats, such as parafluvial and orthofluvial springbrooks. Yet, as John and I were talking, we realized that he and I have very different notions of what constitutes a springbrook habitat. Before we could advance any further, we would need to define our term; otherwise, we would have to stratify our study area and describe the Hoh system very differently.

As scientists, fish managers, and conservationists, we need a common language. Ultimately, our task is to produce precise lab work and accurate fieldwork. Both must be replicable in order to be tested and validated, but without standardized protocols, we will continue to collect data that reflect variations in methodology and cannot be stitched together across project or jurisdictional boundaries. Without common definitions, shared objectives, and standardized protocols, our data will continue to be localized, incomparable between researchers and localities, and, ultimately, fixed in scale.

The ecosystem approach we are moving towards in salmon study is an evolution. Essentially, with the development of ecosystem and North Pacific perspectives on salmon conservation, many methodologies still utilize only site-based studies, capturing only small fragments of larger river basin ecosystems. Unfortunately, although there is increasing collaboration across stakeholders, there remains relatively little collaboration across federal agencies, state agencies, and universities. Seldom are there good ecosystem descriptions with consistent methodology at regional and larger scales.

Certainly we have difficulty speaking across nations, but we also stumble when we attempt to reconcile data at different scales or when comparing data collected in different river basins using similar monitoring techniques. Specialties present yet another level of complexity; perhaps a scientist is well practiced in snorkeling but may not have a thorough grasp of other monitoring techniques conducted in the same watershed for the same purpose.

We are at a crossroads in salmon study. We are advancing concepts by applying an ecosystem approach to our research; at the same time we are making exciting technical advancements through geographic information systems, which enable us to work at multiple scales and provide an opportunity to aggregate data in a spatially explicit manner.

Our objectives as scientists studying a migratory species at multiple scales require us to evolve our techniques as practitioners. We are in a similar place the scientists were in 250 years ago, when our predecessors acknowledged the need for a common language to study salmon. It is not that long ago, after all, that George Wilhelm Steller, in the 1740s, phonetically transcribed the Koryak name for salmon into his German manuscript. Over the next 50 years, his book was translated into French, English, and Russian; it was not until 1792 that Steller's book was Latinized by Dr. Johann Julius Walbaum to conform to scientific naming conventions. The Linnean classification system was needed to identify organisms across the world, but there was no extension of that concept to ecological terminology and methodology.

It is time to develop a shared vision and standardize techniques and to create new methodologies that are scaleable and comparable. As we begin to craft these techniques, we first need to ask some hard questions:

- How do we bring the methodology up to the conceptual level?
- How do we choose the best method to achieve our objective?
- How can we ensure that we can cross-validate findings?
- How can we collect data that can be used at multiple scales?

- How can we ensure that our data can be seamlessly used by other scientists using different methods?
- And conversely, how can we use aggregate data from our colleagues to enhance our own work?

If, for example, our objective is species composition and habitat assessment, then try snorkel surveys. Is the water too turbid? Try electrofishing. This also requires weighing the benefits of accurate size measurements against stress delivered to fish and the food web from electroshocking.

Moving in the direction of a common language and standardized data collection methods, we will be able to make our work more replicable and therefore more credible and useful in the policy process. We will be able to compile the best available information, increasing the precision of field-derived data. This in turn will strengthen the application of theory to real-life circumstances. We will fulfill the needs of multiple stakeholders on a more neutral footing. We will be able to work at multiple spatial scales and compare data from different studies and different rivers around the world, which would open up the types of analysis that could be done on existing data sets (e.g., comparisons from different teams within and between river basins).

These are worthy goals, but here is an even more convincing one: our resources as field biologists and managers are becoming increasingly limited, as is the time remaining to make a difference in the conservation and preservation of our study subjects. We need to optimize our efforts and maximize the use of data that is now available and increase the quality of data that will become available. By giving managers valid information at a broader scale, we can make the leap to using reliable research to inform policy and to link science, management, and conservation.

The protocols in this book have been developed, fine-tuned, tailored, and perfected by field biologists and other practitioners. Let us take these efforts to another level and share this vision. Common language and standardized protocols will improve our work as individual researchers, fishery and habitat managers, and stewards of the river basins we call home.

The Role of Sample Surveys: Why Should Practitioners Consider Using a Statistical Sampling Design?

Donald L. Stevens, Jr., David P. Larsen, and Anthony R. Olsen

Introduction

The primary purpose of this book is to describe the great variety of field sampling protocols for determining the abundance, distribution, and productivity of salmonid populations, especially in stream and river networks. These protocols guide the field practitioner in the selection of appropriate methods to collect fish once the sampling locations have been determined. Equally important is the selection of the locations where fish are to be collected, especially when it is impractical to conduct a census by which all the fish are counted or when information is required for all locations on the stream network. Statisticians sometimes distinguish these two aspects as the sampling or survey design (Where should I collect the fish?) and the response design (How should I collect the fish?) (Stevens and Urquhart 2000).

This chapter provides a small amount of balance across these two critical parts of developing a program to monitor salmonid populations by describing some important components of survey designs relevant to the estimation of the abundance, distribution, and productivity of salmonid populations. A variety of statistical books (e.g., Särndal 1978; Cochran 1987; Thompson 1992; Lohr 1999) cover many of the aspects of survey designs in great detail. Some information on environmental sampling is provided by Gilbert (1987), Olsen et al. (1999), and Stehman and Overton (1994). Stevens and Urquhart (2000) discuss technical issues that arise with response designs when conducting a survey in environmental settings. Two recent books provide some insights useful to fisheries workers; chapter 7 of Thompson et al. (1998) is devoted to statistical sampling of fishes, and Thompson (2004) focuses on the topic of sampling for rare or elusive species. Finally, a forthcoming American Fisheries Society book, Analysis and Interpretation of Freshwater Fisheries Data, edited by Michael Brown and Christopher Guy, will greatly aid fisheries workers in freshwater systems of North America.

We strongly recommend that survey design statisticians be involved as part of the planning team from the beginning in the development of a monitoring project. As members of the team, they have a contribution to make in setting goals that can be objectively evaluated through the development of survey designs and analytical procedures that are consistent with the project goals. More often than not, however, monitoring proceeds without input from a statistician until the analytical phase, when a statistician is often asked to assist in the analysis. By this point, it is often too late; objectives may not have been clearly defined, or the monitoring plan may not match the stated objectives. We liken the circumstances to driving to a new destination without a map: we may get there with the help of a good sense of direction and lucky hunches, but we would certainly save time, expense, and angst if we first invest in a spatial plan before setting out on a trip.

Why use statistical surveys?

Regardless of the field of inquiry—whether it is fisheries or human health, economic vitality or agricultural resources, or human population demographics or labor statistics—an accurate representation of the resource is a necessity if a census cannot be conducted. In all these fields, there is a long history in the development and application of statistical surveys to meet analytical needs. An instructive book that covers the history of election polling—Survey Research in the United States: Roots and Emergence 1890–1960 (Converse 1987)—illustrates the evolution of sampling techniques, from error-prone judgmental selection techniques and targeted sampling to sophisticated statistical surveys currently used.

Some of the important objectives of sampling salmon populations, driven by various agency legislative mandates (e.g., ESA) and management objectives, include the estimation of the number of fish of a particular species within a population, metapopulation, or other demographic unit across that unit's spatial domain and whether these numbers are changing over time. This type of objective is shortened to a "status and trends" estimation. Knowing the spatial distribution or spatial structure of these populations is also important: Are they clustered in one part of their domain or more or less evenly distributed? Is that spatial structure changing over time? Other questions include the following: What proportion of the fish population is of hatchery origin? What proportion is wild? What is the age structure of the population? How large is the breeding population? To what extent do the fish stray from their natal domains? In some instances (e.g., migratory populations), some of these questions can be answered if the fish can be counted accurately (i.e., through a census) as they pass a particular location; however, many of the objectives cannot be achieved in this way, especially if the fish do not migrate. If a census is not feasible, then we must realize our objective by extrapolating from the characteristics of a sample and applying them to the characteristics of the population.

Commonly used techniques for selecting a sample include convenience, representative, model-based, and probabilistic. The first two have been widely used in fisheries and environmental management in general but have some substantial deficiencies. A convenience sample is just that: there is no particular reason for selecting a site other than because it was easy to do so. The relationship between sample data and population characteristics is unknown, and there is no reasonable or defendable basis for extrapolating data from the sample to population characteristics. Such data are often inappropriately analyzed using common statistical tools. A representative sample is most often based on professional judgment founded on an informal synthesis of the investigator's experience. One problematic issue is that a site representative of one variable is not necessarily representative for any other variable; another is that if the sites truly are representative of central tendency, then the extremes are suppressed. A major weakness of this technique is that humans fare poorly when integrating new data due to the existence of prior conceptions; this theory is supported by many experiments in cognitive psychology.

Model-based sample selection uses prior knowledge and theoretical population characteristics to choose sample sites. The model defines the relationship of the sample to the target population and provides a prescription for extrapolation from the sample to the population. A significant advantage of model-based procedures is that they maximize the leverage of data: strong inferences can be made with relatively little information. Although there is a minimal need for data, model-based procedures are not necessarily easy or quick to apply. Construction and calibration of an appropriate model can be very time consuming and expensive. The inference is based on the assumed completeness of knowledge and applicability of the model, and there is often no direct way to verify model assumptions. Without demonstrated reliability based on extensive field data, model results may not be viewed with confidence, and the usefulness of the model in fisheries management, especially in controversial situations, is limited.

The final way to select a sample is by using probability-based methods. Probabilistic sampling has a number of advantages as compared to other sample designs. Survey methodology furnishes a rich array of ready-made inference tools to estimate population characteristics, with known, quantified certainty. Prior knowledge and theoretical understanding can be incorporated, both to focus the design and to sharpen the analysis. Without a census, a statistical survey with the incorporation of probability sampling is the only way to assure the selection of a representative sample from which can be drawn unbiased conclusions about the population as a whole. In fisheries an unbiased estimate of the number of fish is critical; if the estimates are biased, consequences (e.g., species extinction) can be expensive or unacceptable.

An example taken from Oregon Department of Fish and Wildlife's (ODFW) coastal coho salmon *Oncorhynchus kisutch* population monitoring program illustrates the importance of obtaining a representative sample by applying probability site selection methods (Jacobs and Nickelson 1999). Beginning in the late 1940s, salmon spawning runs were monitored with standard spawning surveys at index sites to evaluate escapement—first, past the commercial net fishery; later, from offshore troll and sport fisheries. These index sites were likely selected in better-than-average spawning locations in Oregon's coastal watersheds. By 1980, ODFW realized the need for more exact data to establish population levels and develop harvest regulations. During the 1990s, ODFW incorporated random site selection into the evaluation of population sizes; the agency has continued to use random sample surveys since that time. A side-by-side comparison was made between coho spawner densities at the standard index sites and densities estimated from the random surveys; results indicated that estimates derived from the random surveys were, on average, 27% of the densities derived from the index surveys. Index surveys are used frequently in environmental monitoring, and there is often strong reluctance to shift towards the use of statistical surveys, even in the presence of information that reveals the biases of index surveys. Clearly, as illustrated by this example, results from judgmental or convenience sampling can be severely biased, and unless the bias is corrected, poor management decisions could be implemented.

The key point here is that adoption of statistical surveys allows unbiased, rigorous estimates of many of the important aspects of salmon populations that interest resource managers, policy makers, research scientists, and, perhaps most importantly, the public. Moreover, the estimates have a known, quantified level of uncertainty. Effective, cost-efficient management requires sound data, and appropriately designed statistical surveys can facilitate the integration of data obtained from multiple projects and agencies to provide additive statistical power.

Moreover, surveys can allow information gathered in a particular setting to be combined with data collected under other settings. This capability is especially important because many different agencies have the same objectives and often monitor the same salmonid populations. Furthermore, because salmonid populations have ranges that cross jurisdictional boundaries, comparable data are imperative for sound management. For example, in 2002, the Bonneville Power Authority wanted to knit together discrete salmonid monitoring efforts conducted in the Columbia River basin to create a whole picture of salmonid population status from the sum of its parts; however, in part because statistically reliable sampling designs were rarely used, the data could not be combined to create the bigger picture. Information pertained only to the setting in which it was collected.

Data collected in the field are only as good as the sites selected for monitoring; much like building on a solid foundation, if the frame of our inquiry is correct, our data will stand up. If, however, we build on a shaky foundation—or worse, on no foundation—the information we collect will have little or no meaning.

The importance of establishing survey objectives

A clear statement of objectives is essential to developing a sampling design. Working out the objectives in sufficient detail to guide the development of a sampling design can be a lengthy process. The process begins with conceptual questions (e.g., what are the status and trends of coho salmon on the Oregon coast?). The conceptual question is necessary, but it is neither precise nor detailed enough to design a sample. The final objectives statement must include an operational definition of the target population as well as specifications about which characteristics of the population are to be estimated, what measurements are to be made at the sample sites, and what level of precision is required.

The objectives drive the sampling design because the design is created to satisfy the objectives. Frequently, however, the process is iterative as the objectives are refined subject to sampling feasibility. For example, the nominal objective may be to estimate the size of a native trout population in a basin. If a portion of the basin is too difficult or impossible to sample, we may have to settle for a more modest objective, such as estimating the size of a population in the accessible part of the basin.

Characteristics of survey designs for fisheries

The use of statistically designed sample surveys (or probability surveys), in conjunction with appropriate field sampling protocols, allows for robust estimation of salmonid numbers and for changes in abundance over time, spatial structure, and various other aspects of population structure, as well as for an estimate of sampling precision or uncertainty. In what follows, we will introduce several common approaches for developing survey designs, such as simple random sampling, stratified random sampling, and systematic sampling; we will also describe some of their shortcomings with respect to sampling fish in stream/river networks.

At the same time, we will advocate the use of spatially balanced sampling. In particular, we encourage the use of the generalized random-tessellation stratified (GRTS) design (Stevens 2003; Stevens and Olsen 2004), which overcomes many of the shortcomings of other survey designs. GRTS is described later in this essay. We will also describe the concept of a master sample (Yates 1953) and its potential usefulness to facilitate the integration of multiple monitoring programs across a region such as the Columbia River basin.

The development and implementation of sample surveys applied to stream/river networks presents a variety of practical challenges that must be accommodated by the chosen survey design. GRTS is flexible enough to overcome many of the challenges that the other design approaches cannot. For example, the design should accommodate the following practical issues:

- Spatial relationships among sites. The target population exists in a spatial matrix, and spatial relationships in the population are critical, both to our understanding and to sample design. Sites near one another tend to be similar because they tend to share a number of characteristics such as substrate, climate, topography, and natural and anthropogenic stressors. These spatial relationships lead to patterns in the response, such as gradients, patches, or periodicities. Good survey design takes advantage of the patterns.
- Accurate and relevant frame representation. The representation of the stream/river network (or the frame), such as a digital file, from which to draw the sample is rarely an accurate portrayal of the target population's domain. The frame may include stream segments that are not in the target population (e.g., the segment may be dry). Alternatively, the frame may exclude segments that are in the target population. Sites selected from the frame sometimes cannot be sampled because, for example, access is denied by reluctant landowners or because the site is too dangerous to reach. A good survey design should have some means of addressing unreliable frame material and access difficulty.
- Ability to focus on subpopulations as well as the overall population. In many assessments, subsets of the population will be of particular significance. The significance may arise from ecological considerations, genetics, economic importance, environmental stressor levels, scientific interest, or political pressure. Whatever the source, a survey design must be able to focus on selected subpopulations as well as overall population characteristics.
- Ability to evolve goals and objectives. For monitoring plans that are anticipated to persist for several years, it is almost certain that the goals and objectives of the program will evolve. The survey design must have a substantial amount of flexibility to respond to such changes while maintaining continuity.
- Consideration of time and seasonal constraints. Field crews run out of time and cannot sample all sites designated to be sampled. Some species are distributed in disconnected patches across stream/river networks, for example, and only occupy headwaters during a portion of the year. In some cases, it might be important to sample multiple species simultaneously (as a cost-effective measure), even though those species domains might only partially overlap.

The approach advocated in this chapter is designed to accommodate many of these practical aspects of applying survey designs to stream/river networks.

Survey designs for fisheries

The simplest probability-based or statistical approach to the selection of locations in stream/river networks at which to sample is simple random sampling (SRS; see Figure 1). In SRS, every location in the target network domain is given an equal opportunity to be selected in the sample. This equal opportunity meets the basic statistical criterion that every location has a known, nonzero chance of being included in the sample. No location is selected for convenience or on the basis of expert opinion about a valid reason to select a particular site. Locations selected for specific reasons, such as convenience or because a site has been monitored for a long time or is thought to be representative, do not meet this most basic of statistical requirements.

SRS can be applied to a stream network by splitting the network into segments, arranging the segments in a list, and then selecting segments to sample from the list. The segments can be fixed length, defined by starting at the mouth of the network and working up. They can also be defined by physical features of the network, such as confluences, riffles, or pools. Alternatively, points on the stream network can be picked by viewing the network as a continuum of points. The network can be mapped onto a line, and the SRS sample from the line can be mapped back to the network. In either case, the important characteristic is that each point or segment is chosen independently of any other point or segment. It is then easy to augment the design to account for missing or inaccessible sites or to refocus the sample to account for evolving goals. But SRS makes no attempt to account for spatial pattern in the response and thus will likely be an inefficient design.

At times, knowledge about the distribution of fish can be used to stratify the sampling to devote proportionally greater sampling to some strata. Stratified random sampling does allow greater precision of estimates if the information used to create the strata is correct. For example, we may choose to sample segments with a history of high fish density more intensively than segments with historically low density. If that turns out to be the case, then the estimate of total number of fish will be more precise. In many cases, what we think we know about the distribution and abundance of a species or population across its domain is often erroneous. A consequence of misclassification of the domain into strata is that precision can be lower than it would be if simple random sampling had been used. In any case, the basic rule regarding equal opportunity applies within strata: every location within a stratum is equally likely to be selected. Stratification is also used to ensure that estimates can be made for subpopulations of special interest.

One of the primary disadvantages of SRS and stratified random sampling is that the spatial pattern of sites can be clustered, leaving gaps in parts of the network (see Figure 1). Our visual expectation of a random sample is that the distribution of points will be approximately evenly spread across the relevant domain (or across each stratum within the domain); often, however, this is not the case. Completely random processes such as SRS are much more variable than is commonly thought. Spatial simple random samples exhibit apparent patterns in clusters and voids. That is, the sample exhibits a certain amount of clumping. Spatial stratification with a low number of points per stratum (less than five) can be used to increase spatial regularity and reduce clumping. This comes at a price. Small strata can be difficult to identify or describe, and the effect of missing data is magnified, contradicting the intended increased precision of estimates that could be derived from well-defined small strata.

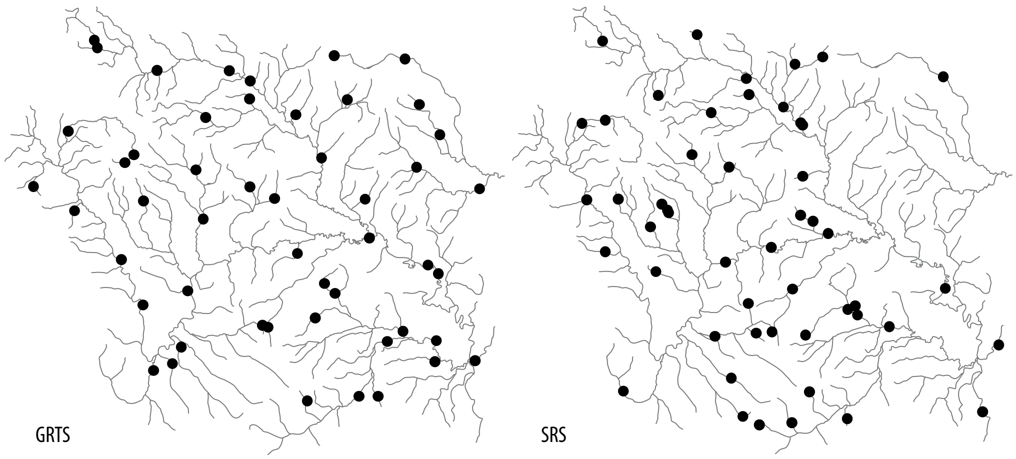

FIGURE 1. — These panels contrast the spatial distribution of sample points established with generalized randon-tessellation stratefied design (GRTS) and a simple random sample (SRS). Notice the relative absence of points in the upper right portion of the network for the SRS sample.

An overriding characteristic of fish populations is that the distribution of fish over their domain often has spatial pattern; parts of the domain might have relatively high densities, and parts low densities. An efficient survey design will utilize this characteristic through spatial balance in the selection of monitoring locations, resulting in increased precision of estimates. Even if we do not know what the pattern is, we can take advantage of its potential existence by ensuring that our sample has spatial balance (i.e., the sample is more or less regularly distributed over the domain). One way of creating an even spatial coverage (or spatial regularity) is to use systematic sampling with a random start. In two dimensions, a systematic sample could be a set of points at the centers or intersections of a square grid or at the intersections of a triangular grid. The known chance of inclusion rule is met by locating the grid with a random start point. For stream networks, a systematic design would look like a set of evenly spaced points across the stream network of interest. But there are disadvantages to systematic sampling. The most severe is that, by accident, the systematic grid might align with the natural feature being investigated. For example, suppose the quality of lakes in Oregon and Washington were being investigated. One set of lakes falls along the north–south Cascade Range. It is conceivable that the random start for the systematic grid aligns to miss this set of high elevation lakes or that the systematic grid preferentially selects these lakes. It is not so easy to identify examples of this type of systematic alignment for stream networks, except for the geomorphic pattern of pools and riffles that roughly occurs in a predictable pattern. Furthermore, poor frame information and missing or inaccessible sites break up the regularity of the initial sample, and it is difficult to add sites while maintaining the regularity.

Adaptive sampling

The distribution of fish sometimes requires adapting the specified design based on what is found at the sites selected. For example, some fish species are distributed in patches across their range. A desirable feature of a sampling design would be to allow increased sampling effort in the vicinity of sites where fish are found when the design selected sites are visited. The concept of adaptive sampling meets this need (S. K. Thompson 1990, 1991, 1992; W. L. Thompson 2004). When a site selected in the original design is visited and no fish are found using the specified sampling protocol, no further sampling occurs in the vicinity of the site; however, if fish are collected at the site, then additional sampling occurs upstream and downstream of the site, following specified rules regarding the length of stream sampled. If no fish are found, then sampling stops. If fish are found, then sampling continues, again following the sampling protocol. Sampling continues until a specified stopping rule is reached. If fish distribution is patchy, the use of adaptive sampling allows for improved estimates of the abundance and spatial structure of target species.

Multistage sampling

Efficient selection of a probability sample requires a good representation of the resource to be sampled. For example, the frame must be a faithful representation of the stream network or estuarine resource to be sampled; however, if the frame is an inadequate representation of the resource, two-stage or multistage sampling can be used. In the first stage, a probability sample of the potential target resource is obtained; then this sample is evaluated to determine whether it is indeed part of the target resource. Ideally, this determination would be reasonably quick and inexpensive, relative to sampling for the relevant indicators. For those first-stage samples that meet the target resource criterion, a second-stage sample, from which the target populations of interest would be measured, could be selected. The concept can be extended to multiple stages if the selection of the final set of sites is efficiently facilitated. Otherwise, the multistage sampling provides no benefit.

The concept of multistage sampling can be extended to include refining a sample even if the frame is an accurate depiction of the resource. Relatively inexpensive measurements could be made on the first-stage sample; then a subset of the first-stage sample could be visited to make more expensive measurements. The first-stage sample could be classified to direct the allocation of the second-stage sample to maximize efficiency with respect to the target estimates. Results from the second stage sample can be extrapolated back to the first stage sample and then to the target resource.

Generalized random-tessellation stratified design

Numerous methods have been developed that adapt and extend the concept of a systematic sample applied to environmental resources to achieve a spatial representation through spatial balance. Stevens and Olsen (2004) reviewed many of these methods, indicating that ". . . all do reasonably well at getting a spatially balanced sample under favorable circumstances, but have difficulties with some aspect of environmental populations." They proposed a solution, via the GRTS design, that builds on and overcomes the difficulties associated with the available designs and creates a spatially well-balanced, random selection of locations on linear (e.g., stream networks) or areal (e.g., estuaries) resources (see Figure 1). GRTS accommodates the following possibilities:

- The resource of interest might exhibit spatial patterns (parts of the domain might have relatively high abundances and other parts low abundances) or patterns might be regular;
- The representation of the resource used to select the sample can be imperfect (e.g., the digital representation of stream networks contains errors; the map of the estuary strata of interest is not perfect);

- The locations can be selected with variable but known likelihood of inclusion in the sample (to achieve flexible stratification);
- Data might not be collected at some sites for a variety of reasons (e.g., site access might be denied, site is too dangerous to visit, field crews run out of time and do not sample all selected sites);
- Sample points can be added to the initial list to accommodate visit denials or to increase sample sizes (points can be added to sample the entire population or only selected subpopulations or subdomains); and
- Multistage or adaptive sampling might be necessary.

The current version of GRTS creates an ordered list of sites such that each successive site on the list maintains the spatial balance of the full set of sites in the sample. The significance of this property is that an investigator can work down the ordered list to achieve his/her target sample size. If a site is not accessible for some reason, that site is skipped (keeping track of the reason that the site is skipped). This process continues down the list until the requisite number of sites has been sampled. If by some fortuitous circumstance a greater number of sites can be sampled than originally planned (perhaps due to a budget surplus), continuing down the list of ordered sites allows the investigator to increase the sample size, yet maintain the spatial balance of the full set of sites.

A Web site has been established that describes GRTS and many example applications (www.epa.gov/nheerl/arm). Algorithms both for creating GRTS designs and analyzing resulting data are available through the Web site. The algorithms are written in the statistical language "R," a freely available statistical software package (R Development Team 2006); the Web site contains information on how to download and install R.

The role of statistical surveys in combining data from different surveys

Among the many advantages of well-designed statistical surveys is that they can allow the integration or rolling-up of data across several surveys that might have been conducted for disparate purposes but monitored the same indicators. This integration is facilitated if several design rules are followed. One is that the surveys use a common representation (or frame) of the target resource from which the set of sites is selected or representations that can be matched. For stream surveys, a common representation of the stream network is the U.S. Geological Survey digital stream network at the 1:100,000 scale. A second design rule is that randomization is used in the selection of sites with the result that the probability that any particular site is selected is known. A third design rule is that the part of the frame that is to be monitored is clearly described. For example, one stream survey might only be interested in headwater or first-order streams. Another might be interested in all first- through third-order streams. Data from the two surveys could be combined if the above rules were followed by recalculating the site inclusion probabilities for the overlapping parts of the surveys. The investigator interested in headwater streams could build in the first-order stream data from the second survey. The investigator interested in the first- through third-order streams could add the first-order stream data from the first survey to the data set. In both cases, the effect is to increase sample sizes—and therefore improve precision of the resulting estimates.

Status and trends and the use of rotating panel designs

The term "status" implies a snapshot of condition during a specified time interval (e.g., how many adult spawners were present in Oregon's coastal streams during the 2006 spawning season, or what was the habitat condition during the years from 2004 to 2006). Precision of status estimates depends in part on sample size: the more sites visited, the greater the precision of the estimate.

Evaluating trends implies study of change over some time interval. If spatial patterns in the response tend to persist through time, then revisiting the same sites during the time interval of interest is optimal for trend detection. Estimating trend from the revisited sites eliminates some site-specific components of variation, so the resulting trend estimates are more precise. Balancing the need for more distinct sites for status estimation and revisiting individual sites for trend detection creates a design challenge that is resolved by the implementation of panel designs (Kish 1987; Skalski 1990; McDonald 2003). A panel consists of a set of sites visited on a specified pattern over time. For example, one kind of panel design might consist of a set of sites visited every year (annual panel), a set of sites visited every 3 years beginning with year 1 (year-1 panel), a set of sites visited every 3 years, beginning with year 2 (year-2 panel), and a set of sites visited every 3 years, beginning in year 3 (year-3 panel). The sites can be selected from a GRTS sample list (e.g., the first 25 sites on the list could be allocated to the annual panel, the second 25 to the year-1 panel, and so forth). Selected this way, each panel is a spatially balanced sample, with the result that each year has a spatially balanced sample consisting of 50 sites.

Although panel designs sacrifice some trend detection capability by devoting more effort to status estimation, their trend detection sensitivity tends to catch up to a strictly annual visit design after three cycles (Urquhart et al. 1993, 1998; Urquhart and Kincaid 1999). In the previous example, after 9 years, the trend detection capability of the panel design would be approximately the same as if 50 sites were visited every year. However, with the panel design, a total of 100 sites are visited, improving the estimates of status. Clearly, in designing surveys investigators must establish priorities for objectives. One design is not optimum for all purposes.

Master sample concept

When multiple organizations agree to use the same field sample protocols (i.e., the same response design) to conduct studies within the same geographic region (e.g., the Pacific Northwest), the next natural question is whether the individual designs can be integrated so that they can take advantage of sampling others have done. The concept of a master sample (Yates 1953) is an answer to this question. The basic idea is to select an equal probability sample over the entire geographic region with a sufficiently large sample size so that the sample size requirements for almost all studies conducted within the region will be satisfied. Such a dense, spatially balanced, ordered list of sites can be selected using the GRTS site selection algorithm.

For example, a set of sites covering the coastal stream networks in Oregon, spaced on average 1 km apart, could be selected with GRTS. This master sample can be classified in a variety of ways such that spatial balance is maintained if the original order of sites is maintained within the class. A target sample size can be selected by proceeding down the ordered list within the selected class. One investigator might be interested in a broad regional survey consisting of 100 sites across the full coastal stream network. Selecting the first 100 sites from the master sample covering that domain provides that investigator with a spatially balanced sample. A second investigator might be interested in sampling only part of that domain—the Nestucca watershed, for example—with a sample size of 50. Selecting the first 50 sites from the part of master list of sites in the Nestucca provides the second investigator with a spatially balanced sample. If both investigators measure the same stream attributes, their data can be easily combined in a statistically sound way. The technical details involve recalculating each site's likelihood of inclusion when the studies are combined.

A master sample for a broad region can serve multiple purposes. As outlined above, it can be used as the basis for designing surveys whose results can be rolled up in a statistically sound way to create a whole picture of status or trend from the sum of its parts, extending the utility of data collected for individual purposes beyond the setting in which they was collected. A master sample is also useful as a tool by which investigators can easily explore allocation of fixed sampling effort (e.g., can only afford to sample 200 sites) in different ways before settling on an optimal allocation. Are a sufficient number of samples allocated to relevant strata? Is the target domain reasonably covered? Numerous alternative designs can easily be created to explore and evaluate alternative designs yet maintain the basic survey design principles of spatial balance and randomization in site selection.

Perhaps a more far-reaching use of a master sample is to facilitate the integration of numerous monitoring programs across a broad region like the Columbia River basin. Currently there are tens of agencies, including federal, state, tribal, and private agencies, conducting monitoring programs within the Columbia River basin. Although these agencies have their own specific goals and objectives, they often are interested in similar riverine and riparian attributes, including estimating various fish population sizes and their physical and chemical habitat. Even though their spatial scales might differ (for example, across the entire Columbia River basin or statewide to the much finer scale of a small watershed), the responsible agencies might benefit by combining data from the other monitoring efforts that are conducted in their domains or at least by knowing where other agencies might be conducting fieldwork. Central housing of a master sample could serve as a design kiosk by which an agency requests a set of sites to meet its specific design requirements. In return, the agency could receive a list of sites along with an indication of which sites are already included in other agencies' monitoring plans. Duplication of effort could be avoided, integration of data derived by common protocols could be facilitated, site selection would follow design principles allowing valid statistical combination of data, and communication among agencies could increase. The next step could be institutional adoption of the concept of a master sample and the formation of a center that provides survey design assistance and coordination in the delivery of survey sites.

Conclusion

Determining the number of fish in streams and rivers or estuaries, their spatial patterns, and how those numbers and patterns change over time poses numerous challenges—not the least of which is determining where to collect fish. Although counting all the fish might be a desirable goal, it is only occasionally feasible;

therefore, statistically sound methods are necessary to estimate fish numbers and patterns. These methods fall into the broad category of statistical sample surveys.

Survey sampling has a long history in a great variety of areas and is built on a strong theoretical foundation. In this chapter, we argue for the incorporation of survey sampling techniques to monitor salmonid populations if a census cannot be conducted. We briefly review the variety of design approaches available, including simple and stratified random designs and systematic designs, and their shortcomings for environmental sampling. We also describe a newly developed, flexible approach that overcomes these shortcomings. We describe the concept of a master sample and suggest its role as a central organizing principle to facilitate integrated monitoring across the numerous agencies now conducting salmonid population monitoring.

Acknowledgments
Dana Foley motivated this essay and contributed regular editorial and formatting suggestions. David H. Johnson provided additional comments. Jennifer O'Neal and Jeff Rodgers reviewed the manuscript. We are grateful to all four for their rapid and timely assistance. This document was prepared at the U.S. EPA National Health and Environmental Effects Research Laboratory, Western Ecology Division, in Corvallis, Oregon, through Cooperative Agreement CR831682-01 to Oregon State University. It has been subjected to the agency's peer and administrative review and approved for publication. Mention of trade names or commercial products does not constitute endorsement for use.

Literature Cited

Cochran, W. G. 1987. Sampling techniques. 3rd edition. John Wiley & Sons, New York.

Converse, J. M. 1987. Survey research in the United States: roots and emergence 1890–1960. University of California Press, Berkeley.

Gilbert, R. O. 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold, New York.

Jacobs, S. E., and T. E. Nickelson. 1998. Use of stratified random sampling to estimate the abundance of Oregon coastal coho salmon. Oregon Department of Fish and Wildlife: Final Report Fish Research Project # F-145-R-09, Portland.

Kish, L. 1987. Statistical design for research. John Wiley & Sons, New York.

Lohr, S. L. 1999. Sampling: design and analysis. Duxbury Press, Pacific Grove, California.

McDonald, T. 2003. Review of environmental monitoring methods: survey designs. Environmental Monitoring and Assessment 85(3):277–292.

Olsen, A. R., J. Sedransk, E. Edwards, C. A. Gotway, and W. Liggett. 1999. Statistical issues for monitoring ecological and natural resources in the United States. Environmental Monitoring and Assessment 54:1–45.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Särndal, C. 1978. Design-based and model-based inference for survey sampling. Scandinavian Journal of Statistics 5:27–52.

Skalski, J. R. 1990. A design for long-term status and trends monitoring. Journal of Environmental Management 30:139–144.

Stehman, S. V., and W. S. Overton. 1994. Environmental sampling and monitoring. Pages 263–306 in G. P. Patil and C. R. Rao, editors. Handbook of statistics, volume 12. Elsevier Science, Amsterdam, The Netherlands.

Stevens, D. L., Jr., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99:262–278.

Stevens, D. L., Jr. 2003 and A. R. Olsen. Variance estimation for spatially balanced samples of environmental resources. Environmetrics 14:593–610.

Stevens, D. L., Jr., and N. S. Urquhart. 2000. Response designs and support regions in sampling continuous domains. Environmetrics 11:13–41.

Thompson, S. K. 1990. Adaptive cluster sampling. Journal of American Statistical Association 85:1050–1059.

Thompson, S. K. 1991. Stratified adaptive cluster sampling. Biometrika 78:389–397.

Thompson, S. K. 1992. Sampling. John Wiley & Sons, New York.

Thompson, W. L. 2004. Sampling rare or elusive species. Island Press, Washington, D.C.

Thompson, W. L., G. C. White, and C. Gowan. 1998. Monitoring vertebrate populations. Academic Press, Inc., San Diego, California.

Urquhart, N. S., and T. M. Kincaid. 1999. Designs for detecting trend from repeated surveys of ecological resources. Journal of Agricultural, Biological, and Environmental Statistics 4(4):404–414.

Urquhart, N. S., W. S. Overton, and D. S. Birkes 1993. Comparing sampling designs for monitoring ecological status and trends: Impact of temporal patterns. Pages 71–85 in V. Barnet and K. F. Turkman, editors. Statistics for the environment. John Wiley and Sons, London.

Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications 8:246–257.

Yates, F. 1953. Sampling methods for censuses and surveys. Hafner Publishing Company, New York.

Data Management: From Field Collection to Regional Sharing

This chapter has been written in two parts. Part I covers field data collection and emphasizes working with local-scale, observation-based data; it frames key data concepts and structures that allow local data to be connected subsequently with larger, distributed regional data systems. Part II deals with regional data sharing and focuses on the features of an effective network of regional data centers; the function of these centers is the generation, repository, and dissemination of data as a resource to people and organizations interested in regionwide, national, international, and global natural resource issues.

Introduction

The purpose of collecting any data is to help answer a question! Thus, developing and using protocols may result in a data-rich endeavor; however, this is not a guarantee that the necessary information will be available to answer a local or regional question(s). With the Information Age upon us, technology plays a large and important role in gathering, compiling, and synthesizing data. Today's issues and their complexities may potentially overwhelm resource managers in a sea of data; yet when resource agencies are presented with a concern or issue, managers may find themselves confronting a lack of information. The need to analyze data over time and space today requires an increased use of technologies, including their integration into research and monitoring studies as well as evaluation strategies. Resource managers must understand that data standards and protocols help refine the quality of data being collected, enhance its usability, and clarify its purpose. Without standards and protocols, resource managers will have only disparate data sets that contain various kinds of information to answer increasingly more complex questions at various scales (e.g., site, watershed, subbasin, and basin levels).

To this end, fishery managers face an urgent need to standardize information; postponing it only exacerbates the problems. Managers need high-quality, real-time data that can be shared by others. As our data management capabilities expand in unexpected ways from year to year, practitioners face the tremendous challenge of keeping up with what's available; broadening our horizons to consider new ways to manage data can be daunting, but we have to rise to the challenge. More than ever, emerging technologies are outgrowing old templates. Certainly we have abundant material on data collection at the local level, but our technology permits us to go beyond the local to the regional and global scales. This essay confronts the necessity of collecting field data along with creating designs for regional data structures and explicit management questions; it may also enlighten us as to how we can create international legislation to foster global data management systems.

The protocols described in this book allow us to open the door to advanced data management systems that can have regional and global applications; however, to start, we need to begin with field data. The development of this chapter began in 2004 with a 2-day meeting in Welches, Oregon, convened with the sole purpose of providing guidance for the design of observation-based data (OBD) system, especially for collecting fish-related data. Data management systems must be designed from end to end (i.e., data collection through data entry to database management, report production, and data sharing). To clarify these mechanics, we have framed this chapter in two parts.

Part I: Field Data Collection

Stewart Toshach, Richard J. O'Connor, Thomas A. O'Neil, Bruce A. Patten, Cedric Cooney, Bill Kinney, Paul Huffman, and Frank Young

In developing this guidance for field data collection, we recognize the following assumptions:

- Local data systems that are developed will feature free and open access.
- Data collection programs will have a principal investigator (PI) who understands the business and scientific logic of the effort and that data management support is needed from a data manager (DM).
- The DM for the project knows that the PI is responsible for identifying the program goals, the data collection deliverables, and the data products. The relationship between the PI and the DM is a critical one, in which success depends on excellent communication across different organizational disciplines, as well as through the planning, execution, and documentation stages of the project.
- Data will be shared, which means that the role of data owner and data steward must be identified and that the data itself must be sharable. With respect to these roles the data owner (who may be the PI) gives permission to the data steward (who may be the DM) to provide access to other users of the data based on agreements.
- Mutual benefit to sharing data is realized by encouraging participants to both use resources and contribute data, information, and knowledge.
- Data contributors should have full right to attribution for any uses of their data, information, or knowledge and the right to ensure that the original integrity of their contribution is preserved. Users of the data are expected to comply, in good faith, with terms of uses specified by contributors or data stewards.

Collecting data and developing small, localized databases do not necessarily pose problems, because they exist to meet the user's immediate and near-term needs; however, most of these data collectors and developers have little familiarity with data concepts and structures that would allow their contribution to a larger, distributed application. To deal most effectively with data systems designed to support the retrieval and integration of primary data, let us begin by establishing a local data set that lends itself to a common management schema.

Establishing a common management schema

To maximize the usefulness of data collected in the field, database architecture should incorporate the ability to apply data at various scales, such as connecting local data sets into larger regional systems. But before going regionally or globally, we need to establish a local data structure that can be moved into a data system. To do this, we need to have an overview of the process:

(1) Outline core questions that we would like to ask of a data structure.
(2) Define terms of data being collected.
(3) Create a consistent process for developing an OBD project (specifically, preparing a needs assessment and writing a data collection and management plan).
(4) Determine office and field procedures needed for entering data, along with Quality Assurance/Quality Control (QA/QC) protocols, analysis, reporting, and maintaining data.
(5) Design data forms.
(6) Create data fields (elements) and identify those that are required or recommended.
(7) Categorize fields or elements as either data or mapping.

1. Outline core questions

Developing core questions begins with determining the data fields you will need to help answer a local problem or concern. Remember that local data sets drive the types of questions that can be asked at a regional level. When developing your local data fields, be aware of regional needs and how your data set might fit in to help answer another question. An example of some local data questions that might occur within a watershed are

- What are the fish species?
- Where can I find them?
- What fish are listed as threatened, endangered, or of concern, and where do they occur?
- Which fish are native and which are exotic?
- How are the fish populations or selective stocks of interest faring (e.g., how many are there relative to historic populations sizes)?
- Why is a specific fish population declining or increasing (i.e., what are the limiting factors or main causes affecting the fish population(s) and their habitat(s))?
- How much fish habitat exists?
- Where does the fish habitat occur?
- What is the condition of the fish habitat (e.g., what is the assessment of watershed and marine ecosystem health)?
- Who is conducting the research, monitoring, or a management project, and where is this work occurring? Is there a need for public outreach and education of this work?

At the regional level you would follow a process that incorporates review of and consensus on the local data sets, which by necessity should be general in design to maximize the use of the information with other data sets by capturing key data elements. This does not mean that specific questions should not be listed; rather, it only means that they should be a subset under a general one. In other words, do not start out too narrowly focused, because you may miss important aspects that could help answer the specific question(s).

There are two approaches for determining questions and implementing data capture: (1) list the questions along with the data elements needed to answer them; and (2) look at the data sets within a region and determine what data elements are common, and then develop the questions that can be answered. Realistically, a combination of both approaches is probably most efficient. Both approaches call for agreement on common data elements and definitions so that data elements collected in one area have the same meaning within another part of the region.

To maximize the use of a regional data network based upon local data, the general questions need to be carefully articulated to allow for the eventual collection of as much detailed baseline information as possible. Therefore, the best tactic is to ask general questions that can be refined. As an example of this regional, top-down approach, the National Water Quality Monitoring Council (2004) developed a list of data elements considered important for reporting water quality. These elements were then sent to individual state agencies and organizations, which were encouraged to use the information. Along with the data elements, several suggested approaches for incorporating these data elements into their specific data programs were made. They included:

- at the local level, consider using all the data elements or as many as possible in your next water quality monitoring project or in developing your next water quality database;
- focus initially on categories of water quality data elements (WQDE) that you most want to improve in the near term and progressively expand the data elements included over time; and
- plan to include these WQDE in database modernization or updating; and in combination with other approaches, program field electronic devices for on-site entry of field data to download directly to your database.

2. Define basic terms

Defining terms for the data being collected helps ensure a common language in the creation of OBD systems. Here are some basic terms and definitions regarding OBD systems:

- Observation-based data: information generated during an activity performed by participants in which observations are collected about a subject following a methodology at a location during a period and for a particular purpose
- Activity: what you are doing (e.g., fish counting, habitat survey)
- Participants: the individuals or organizations performing roles associated with the activities (e.g., observer, pilot, data recorder)
- Methodology: how you are performing the activity (e.g., beach seining)
- Location: where you are performing the activity (e.g., port name, universal transverse mercator coordinate, latitude/longitude coordinate)
- Period: when you are performing the activity (e.g., 2004-03-11, 1320–1456, 2004 March 11/12:30 p.m., week 34)
- Purpose: why you are performing the activity (e.g., assessment of stock B, environmental impact study, escapement estimate) (This is usually defined within the project description and may be reflected as a title on a data form.)

- Observations: the details about a subject composed of characteristics that have values and may have a unit of measure
- Subject: what you are observing (e.g., a fish, mammal, boat, stream reach, rock)
- Characteristics: type of data element/detail (e.g., length, weight, age, color)
- Values: the category or measure of the detail (e.g., 3, blue, 4.859, XZ2)
- Unit of measure: unit used to report the value(s), where appropriate (e.g., centimeters, pieces, grams)

3. Create a consistent process

The more transparent and consistent the process is, the more reliable the data. To create this reliability for an OBD project, we recommend that the DM and PI prepare a needs assessment, write a data collection and management plan, review the field and office setup, and decide on the methods for data recording, data entry and quality review, data analysis, and data reporting. the steps to estalibsh these are

Planning

Creating a needs assessment: identify and document

- the people involved;
- data management budget;
- expected dates of data management deliverables;
- data outputs needed;
- activities to be done;
- what data will be collected;
- what data will be recorded (see *Data recording* below for guidelines);
- whether existing protocols and legislation are applicable (e.g., are there any data collection or reporting standards that must be met?);
- needed quality assurance/quality control (QA/QC) measures and responsibilities;
- who has user rights to create, read, update, and delete data elements (handle/manipulate data) and when these rights are granted;
- data security needs;
- when and how data and/or analysis results will be made accessible to requesters and for general access;
- data work flow;
- whether existing software and hardware are adequate;
- equipment needed to support data recording and storage (e.g., Personal Digital Assistant (PDA) or other equipment) and supplies;
- responsibility for completing the needs assessment document.

Writing a data collection and management plan

The PI will now be able to write a plan that can deliver the identified and documented needs. The data collection and management plan should identify the people responsible for the different tasks and what they will do. Once completed, submit this plan along with your documented needs for review and approval by the appropriate resource program and information systems staff in your organization.

4. Determine office and field procedures for data entry and handling

Field considerations

- Develop data collection form(s) and test with user group, if necessary.
- Acquire the data collection and management equipment and supplies identified in planning.
- Identify maps, global positioning systems (GPS), or other resources that will assist in accurately recording field location information.
- Train field staff in use of field forms, definitions, equipment, and field techniques.

Office considerations

- Install/develop and test any needed or existing data management hardware and software.
- Train staff in use of hardware and software.

Field and office considerations

- Test data collection system from end to end with representative users.

Data recording

Assumes that the researcher/observer has completed data collection following appropriate methods. (For example, the researcher has used a fyke net to capture salmon fry and is ready to record the observations such as the number of fish captured). Following the guidance document, the researcher should then

- complete data recording tasks, following data recording instructions;
- complete any QA/QC tasks (e.g., read the completed data recording to check for completeness, legibility or obvious errors);
- complete in-field data backup;
- deliver data to the appropriate person in appropriate form.

Data entry and quality review

- Enter data as soon as possible.
- Maintain strict control of file and version names.
- Archive this version of the data by storing a copy in a secured and preferably separate location.
- Perform all specified QA/QC tasks (e.g., double entry of the data can be used to perform data verification while value range checking can be used to perform data validation).

- Report recording discrepancies or errors to data collectors for resolution.
- Fix the records in the working version.
- Archive a copy of the working version.
- Repeat the specified QA/QC tasks (testing, fixing, and archiving) until errors are eliminated.

Data analysis

- Complete data analysis tasks using the current version of the data (e.g., develop the necessary statistical reports, charts, maps, and summary data sets). All derived data sets should be maintained and archived and subject to the same version control guidelines applied to primary data sets.
- If errors in current data set are found, perform all fixes and QA/QC tasks described in the Data Entry and Quality Review section (above).

Data reporting

- Complete a metadata document containing the necessary metadata information, including any limitations of the data. Completing a good metadata record is essential!
- Deliver data and/or analysis results according to formats, protocols, and instructions in the needs assessment document.
- If data or analytical errors are discovered, go back and correct a new copy of the appropriate archived version and label it as the current version to be used, using the current date.
- Audit data collection and management effort (for cyclical data collection projects) for inclusion in future data management updates.

5. Design data collection forms

The PI must ensure that the forms used to record or enter data, whether paper-based or electronic, are readily legible in the environment in which they will be used.

Consideration for data forms

- Appropriate font sizes, colors, or graphics should be used. Where possible, allow the data to be recorded or entered in an intuitive fashion (from top to bottom and from left to right). Electronic forms should use tab-key advance to navigate from one field to the next.
- Some systems may have to be created using more than one software or computer language.
- Clear instructions and definitions should be developed to support the system and be readily available to users. Table 1 gives an example of elements to consider when developing a field data form. If the needed data element will be coded, then it will be necessary to create a table of the codes and the corresponding data elements and definitions. If an electronic system will be used, these tables should be an integral part of that system and should be visible using that system's Help functions.

A growing approach is entering data on Web-enabled forms via PCs and/or wireless devices, due to the ability to provide immediate data validation and storage in a central location for records entered from a distant (field) location. While there are advantages of electronic reporting, there are also many lessons to be learned by developing the early field reporting prototypes using paper-based systems. This allows for the testing of data collection and the logical and mechanical reporting concepts before committing to the more expensive tasks of coding electronic reporting devices and developing databases. While changes to reporting systems are common during development, they can be minimized with careful design and iterative testing.

6. Create the "universe" of relevant data elements

- There is a minimum set of data elements that must be included in OBD collections to ensure that adequate value is derived from the collection effort.
- There are additional data elements that may be included, depending on the nature of the study and other factors.
- All the data elements that will be collected must be identified and described in a formal table.

Table 1 shows a list of the *minimum* set of data elements that must be recorded during OBD collection activities. A minimum set of data elements must be included to ensure adequate value is derived from the collection effort. Additional data elements may be included, depending on the nature of the study and other factors. All the data elements to be collected must be identified and described in a formal table. For each element, provide the data element name, definition, and format. These three columns should be part of a larger data description table (see Table 2), which identifies, categorizes, and maps data elements.

Some data managers may also want to create columns about the needed data element—whether it is required (yes/no) or conditional, based on the provision of other data, and for defined units of measure. The format column can also be used to report the length of characters that will be accepted by the database together with any decimal points. (See Table 2.1 for an example of how salmon spawning stock survey data elements found on an actual field sampling form were listed.)

7. Categorize and map data elements

Each element must be categorized (to ensure that no vital aspects of the study have been overlooked) and mapped (to determine where to record that element) with the help of an appropriately organized data description table.

Categorization

Each data element is assigned to one of the following OBD categories:

- Participants (who?)
- Methodology (what or how?)
- Location (where?)
- Temporal (when?)
- Purpose (why?)
- Subject (of an observation)
- Characteristics (of a subject)

Careful review of these assignments is done to ensure that every element has been assigned. If there are any categories not represented by at least one element, then critical information relevant to your study is not being recorded. Add elements as needed to represent all OBD categories. (See Table 2.2 for an example of salmon spawning stock survey data elements assigned to their appropriate OBD categories. Note that in this example, methodology (how) and purpose (why) categories were not represented by any data elements present on the field form.)

Mapping

All information related to an OBD collection activity must accompany your observations. Each data element needs to be recorded somewhere in your system. Data elements are commonly recorded in one of three locations:

- Project Documentation
- Field data form—header section
- Field data form—detail section

Use the following criteria to determine where to record a data element:

- Project documentation: data element value is fixed throughout the entire study
- Field form—header section: data element value is fixed for a single activity
- Field form—detail section: data element value changes from observation to observation

Aspects of certain data elements can be recorded in more than one place. For example, you may want to enter the time that all sampling started in the form header and enter times of individual observations in the form detail. (See Table 2.3 for an example of salmon spawning stock survey data elements mapped to their appropriate places.)

TABLE 1. — Minimum data for observation-based data collection.

Data Element Name	Data Element Definition	Format	Examples
Activity Name	Brief description of activity	Text	Chinook redd survey project
Name of data collector(s)	Name/s of individuals who are responsible for data collection	Text	Bruce C. Patten; RJO
Date of data collection activity	Calendar date expressed as YYYY/MM/DD	Alphanumeric	2003/05/21*
Time of data collection activity	Time of day expressed in hours, minutes, seconds local time (24-hour clock)	Alphanumeric	06:30:23; 15:15
Location of data collection activity—detailed	Location expressed as a point, line, or polygon with latitude/longitude expressed in decimal degrees (4 decimal places) or equivalent	Alphanumeric	–123.4890, 45.8734
Name of data collection activity location	Name of body of water (lake, stream), place name, or land unit	text	Ward Creek; Lake Washington; Bonneville Dam

Data Element Name	Data Element Definition	Format	Examples
Geographic code for data collection activity location	Standard code for a body of water (LLID) or land unit (fourth field HUC)	Alphanumeric	LLID: 1242059430208 (Ward Creek, Coos Basin, Oregon) HUC: 17030003 (Yakima River—lower and tribs)
The method(s) used during the data collection activity	Identification of the specific method(s) used to collect the data.	Alphanumeric	Method 27; Purse seine; Redd survey protocol (2006)

* Welches Working Group initially recommended MM/DD/YYYY; subsequently the Northwest Environmental Data-Network recommended using the International Organization for Standardization (ISO) format YYYY/MM/DD.

Below are examples of recommended data elements, and if they are needed they should be part of a larger Data Description.

Location:
- Sample/survey length/area:

Subject:
- Species:
- Run (for fish species):
- Sub-run (if applicable):
- Life stage:

Participant(s):
- Agency:

Methodology:
- Sampling method (gear):
- Target (if any):
- Photo available (Y/N):

Characteristics:
- Habitat type:
- Air temperature:
- Water chemistry/quality (temperature, clarity, pH, DO):
- Weather conditions:
- Waterbody physical attributes:

Miscellaneous:
- Page sequence (page _ of _) (especially important for paper records):

TABLE 2. — Example data description table with its three component sections: 1) identifying data elements, 2) categorizing data elements, and 3) mapping data elements.

TABLE 2.1. — Data description table: Identifying data elements (example entries are from the Daily Salmon Spawning Stock Survey Field Form [Duffy 2003]).

Section 1: Identifying Data Elements		
Data element name	**Data element definition**	**Units of measure**
Example		
Stream		
TRS		
Lat./long.		
Quad.		
Drainage		
County		
Starting location		
Ending location		
Feet/miles surveyed		Feet/miles
Date of survey		Daily
Water clarity		Feet
Water temp.		
Weather		
Air temp.		
Time		Hours/minutes
Crew		
Live fish observed		
Chinook adults		pieces
Chinook grilse		pieces
Coho		pieces
Steelhead		pieces
Unknown		pieces
Carcasses examined		
Chinook M Fk L		
Chinook F Fk L		
Coho M Fk L		
Coho F Fk L		
Tag numbers		
Other clips observed		
Skeletons observed		
Chinook		pieces
Coho		pieces
Steelhead		pieces
Unknown		pieces
Redds		pieces
Comments		pieces

Salmon spawning Survey
Not on form
Visual surveys
Purpose

TABLE 2.2 — Data description table: Categorizing data elements (example entries are from the Daily Salmon Spawning Stock Survey Field Form (Duffy 2003).

	Section 2: Categorizing data elements							
	OBD categories							
Data element name	Participants	Methodology (what)	Methodology (how)	Location	Period	Purpose	Subject	Characteristics
Example								
Stream				x				
TRS				x				
Lat./long.				x				
Quad.				x				
Drainage				x				
County				x				
Starting location				x				
Ending location				x				
Feet/miles surveyed								x
Date of survey					x			
Water clarity								x
Water temp.								x
Weather								x
Air temp.								x
Time					x			
Crew	x							
Live fish observed								
Chinook adults								x
Chinook grilse								x
Coho								x
Steelhead								x
Unknown								x
Carcasses examined								
Chinook M Fk L								x
Chinook F Fk L								x

Coho M Fk L					x
Coho F Fk L					x
Tag numbers					x
Other clips observed					x
Skeletons observed					
Chinook					x
Coho					x
Steelhead					x
Unknown					x
Redds observed					x
Comments					x
Salmon spawning survey	X			X	
Not on form					
Visual surveys		X			
Purpose			X		

TABLE 2.3 — Data description table: Mapping data elements (example entries are from the Daily Salmon Spawning Stock Survey Field Form (Duffy 2003).

	Data Mapping		
Data element name	Project documentation	Activity (field form—header)	Observation (field form—detail)
Example			
Stream		X	
TRS		X	
Lat./long.		X	
Quad.		X	
Drainage		X	
County		X	
Starting location		X	
Ending location		X	
Feet/miles surveyed		X	
Date of survey		X	
Water clarity		X	
Water temp.		X	
Weather		X	
Air temp.		X	
Time		X	
Crew		X	
Live fish observed			

Chinook adults			X
Chinook grilse			X
Coho			X
Steelhead			X
Unknown			X
Carcasses examined			
Chinook M Fk L			X
Chinook F Fk L			X
Coho M Fk L			X
Coho F Fk L			X
Tag numbers			X
Other clips observed			X
Skeletons observed			
Chinook			X
Coho			X
Steelhead			X
Unknown			X
Redds			X
Comments			X
Salmon spawning survey		X	
Not on form			
Visual surveys	X		
Purpose	X		

Part II: Regional Data Sharing

Thomas A. O'Neil, Stewart Toshach, and Wayne Luscombe

As we embark on our own monitoring efforts, we need to keep regional data systems in mind; yet we acknowledge that our monitoring concerns will begin at a more local level. The creation of a data management system comes from a question-and-answer process. Regional protocols or standards help refine the quality of data being collected and enhance its usability, as well as clarify its purpose; however; having local or regional protocols or standards is not a substitute for the establishment of an effective regional data network that can answer broader questions about our natural resources. Leadership and an administrative framework are also needed to guide the development of protocols, standards, and guidelines for collection, compilation, and reporting of component data. It is with the establishment and coordination of a network of regional data centers (directed towards the generation, compilation, and dissemination of accurate and complete data) that the major benefit occurs. These regional data centers are responsible for the generation and repository of special databases and for the dissemination of these databases as a resource to people and organizations interested in components of the regionwide natural resource data.

To assist resource managers with this task, this section explores aspects to consider when establishing a regional or global data network. The process of regional data sharing involves establishing an administrative framework, regional data centers, and a data management system, but first and foremost is the need to agree to work together and establish standards, which begins by identifying the principal or core questions that the regional or global data network is expected to answer. These components are considered essential:

- develop consistent data standards and protocols within and across types of monitoring;
- establish close working relationship for data consistency across data sources;
- identify and document specific data needs of the region;
- develop and recommend data collection standards and information to be shared across programs;
- share requirements and results with regional data networking entities;
- test the collection protocols, sampling methods, and data sharing mechanisms;
- implement coordinated solutions;
- embed common analysis capabilities and reporting capacity;
- provide public access sections or linked Web sites.

Establishing an administrative framework

A survey completed in 1995 for African nations identified nearly 100 different environmental information system (EIS) related activities. This number was estimated to be only 10%–20% of the total EIS activities in Africa at that time (Prévost and Gilruth 1997). Prévost and Gilruth's post-UNCED (United Nations Conference on Environment and Development) report estimated that over the prior three decades, support for EIS activities in Africa "from bilateral aid agencies, international organizations and NGOs is probably in the order of US $500 million." This is an example of why standards or protocols by themselves will not advance a regional data network without first establishing leadership and an administrative framework. Because each group may have its own standards and protocols, the need for information is compounded by the potential multiple of uses; hence, some coordination among groups is required.

To ensure its success and sustainability, a number of factors should be considered when designing and developing the framework for a regional data network. These include but are not limited to (1) establishing clear objectives, (2) coordinating initiatives and avoiding duplication of efforts, (3) developing a data strategy, (4) developing a realistic cost estimate, (5) building institutional support for the initiative, and (6) establishing management and technical steering committees. Stewart Toshach from the National Oceanic and Atmospheric Administration (NOAA), and Peter Paquet from the Northwest Power and Conservation Council, who co-lead the Northwest Environmental Data Network, wrote, "The overall goal is to materially and demonstrably improve the quality, quantity, and availability of data and related information in the Columbia Basin."

They then list the initial steps needed to establish an administrative framework:

- Distribute and discuss the draft memorandum of agreement (MOA) with regional stakeholders to gather input with a goal of expanding participation and creating agreement on a common regional MOA.
- Distribute and discuss the draft administrative framework with regional stakeholders with a goal of reaching agreement on an accountable regional administrative mechanism for a regional data network.
- Arrange for the existing project team and coordinating committee to be consolidated into one project team.
- Complete further coordination with other programs serving regional information management needs.
- Make information about the proposal for a regional data network publicly available and continue to solicit public input.
- Proceed to develop a detailed work plan and costs for phase II (to adopt/develop data network protocols and standards).

Developing an administrative framework is a necessary precursor to collecting, analyzing, and disseminating information from a regional data network. This framework is critical for the development of strategies by federal, state, and provincial organizations to evaluate, conserve, and protect our natural resources. One of the foremost steps that needs to occur in any administrative framework is standardizing the use of data elements or "establishing the use of common data elements" among organizations. It reflects agreement on representations, formats, and definitions of common data and metadata and their definitions. Determining standards begins by collaboratively identifying the principal or core questions that the regional or global data network is expected to answer.

Establishing regional data centers

All information is local, and its care and maintenance should remain in local hands. As a regional or global data network wants to be responsive to fish and wildlife needs within a specified area, it must ultimately participate or cooperate with the entities that collect and use natural resource information at the local level. To do this efficiently, identify regional data centers that can be a focal gathering point of pertinent local information needed to answer core questions. FishBase (<www.fishbase.org>) and the Interactive Biodiversity Information System (IBIS, <www.nwhi.org/index/ibis>) are excellent examples of multiple partnerships forged to address regional information needs. Each has more than 40 partners supporting their development. But in addition to establishing collaborative and cooperative working relationships within the region, the principal roles that a regional data center should address are:

Accessing data of local entities and experts

To sustain geographical and environmental information initiatives, building capacity and infrastructure is necessary to take full advantage of the available data. The old adage that the data is only as good as the people collecting it may hold some truth here. The experience of local people, their knowledge of an area, familiarity with the plants and animals, and understanding of customs and accessibility to sites all play an important role in how and what type of data is collected. Local experts may be more willing to bring forward their information when they understand how and for what purpose it would be applied. Thus, if questions arise regarding how and what was collected and why, contacting local sources or experts will help achieve a more reliable data set. Regional data center personnel will need to work collaboratively with local entities and experts because a primary purpose of any regional data center is to collect and disseminate the best information available.

Ensuring data quality

Data quality is always an important consideration because the computer doesn't "care" about the accuracy, reliability, or source of the digital data; it is all treated the same way. For example, many of the early efforts to convert standard topographic series and line maps into digital formats basically only created graphic layers of information with no internal topology, thus rendering them of limited use in geographical analysis. People who collect the data have a better understanding of how to use it and what the limitations are; therefore, a primary charge of any regional data center is to assure the quality of the information that is disseminated by continuing to connect with local experts and, as necessary, have them review the information.

Institutionalizing arrangements

Frequently, initiatives have not been developed with close organizational links to senior-level decision makers. Without these close associations, the initiatives have often been poorly understood by senior managers and therefore have not been given the financial resources or the political support needed for them to succeed. Uninvolved or poorly informed managers are less able to provide direction and guidance concerning the initiative's objectives and functions. The regional data centers need to champion their cause with local stakeholders and organizations to support the program.

Ensuring adequate funding

A successful initiative depends on adequate funding for installation, implementation, data development, and long term maintenance and operation. Many initiatives have been less than successful because adequate funding was not ensured for either the initial implementation or the long-term operation and maintenance. When preparing initial budgets, funding was often identified only for hardware and software components of a system. A general rule often used as a guideline for estimating relative costs is what has become known as the 20–80 rule, which suggests that 20% percent of the total cost is for system hardware and software and 80% is for data development activities, institutional costs, and other operating expenses. Financial support for the needs of regional data centers can be addressed in one of two ways: (1) the administrative framework, and/or (2) developing a demand-driven approach. For the latter method to be successful and sustainable, however, the center has to be able to respond more directly to actual demands and to be more closely and directly integrated into the decision-making processes. A demand-driven approach helps ensure that adequate funding is available because the decision support provided by the system becomes crucial to the decision-making process. The support becomes particularly indispensable if it can prove its economic value by avoiding costly errors in decisions and by helping decision makers arrive at the most cost-effective solutions.

Training

For geographic and environmental information systems to have a significant impact in a region requires more than a sophisticated technical capacity. Human resources must be developed through education and training programs to enable in-region agencies and organizations to take advantage of the information support tools available. In preparing education and training guidelines for sub-Saharan African countries, Van Genderen (1991) suggests that nine different groups should be targeted: decision makers and planners, opinion leaders, managers, resource surveying personnel, technical support staff, research workers, teachers, students, and the general public.

Establishing common data elements

A common data element is a set of information with the same attribute and an agreed-to definition. A set of standard data elements is then a group of common data elements that all have common definitions and are used to record, monitor, or describe situations or conditions associated with a specific activity (NWQMC 2004). Agreeing to use standard or common data elements allows sharing of information, thus enhancing the potential for increased use of recorded data (both spatially and temporally) within and among organizations. (Appendix A offers examples of common data elements from the Darwin Core Program, which is operated by Global Biodiversity Information Facility (GBIF) and was developed to make the world's biodiversity data freely and universally available on the Internet.)

A data strategy should be established with input from the various agencies that are involved in data development and management. It is only through a multiagency effort that a consistent approach can be developed. It is unwise for any one agency to assume independent responsibility for setting data standards associated with information from sectoral agencies such as forestry, agriculture, and transportation. When single agencies attempt to avoid that responsibility, it usually builds resentment and reluctance for other agencies to cooperate in a data standards effort.

Geographic and environmental information initiatives have greatly improved chances of success if the underlying data sets are both technically and thematically compatible. Harding and Wilkinson (1996) suggest that "interoperability of data and software is … a particularly important issue, underpinning many of the other [issues related to successful information development]. Interoperability [applies] at many levels, not only across different database systems but also across hardware platforms, sites and disciplines, and involving interaction between processes and data of many types." Ensuring that databases are interoperable at the outset by establishing protocols for data collection will help avoid additional costs, inconveniences, discrepancies, and duplicate data sets created by different agencies. Data sets that are integration-ready promote and facilitate analysis and help reduce the overall costs of data development.

Well-documented data have standards and thus become more valuable with time, while undocumented data quickly erode. As mentioned, establishing standards allows for data sharing that enables data sets to increase in size, provide greater statistical power, and ultimately have a higher degree of confidence associated with them. Also, because of the multiple needs for information, data sharing increases the likelihood of more accurate and/or comprehensive assessments, because the meaning of each data set and how they fit into a given context are better understood. Last, the individual data set increases in value through use of common data elements because they increase the potential of using the data for purposes other than what was originally intended (NWQMC 2004). This potential can be quickly assessed in the metadata developed for each data set.

Most data collected for fish and wildlife have common elements like point of contact/collector name, date and time, species, and location. Table 1 lists the minimum elements that should be included on OBD collection forms. Of these elements, location is often of keen interest. When dealing with spatial data elements, there are several recent technical factors that help with standardizing data elements. They are the common hardware and software used to capture, record, and display the data. To capture and record data points, lines, or polygons, GPS units (especially handheld ones) have come into favor. It is important to understand GPS and how it can be used, because it is currently the most common way of acquiring a detailed location. GPS units help land, sea, and airborne users locate where they are on earth 24 hours a day by triangulating earth-orbiting satellites; typically, three satellites are needed to obtain a triangulation, but four or more render more accurate results. The GPS unit is actually a receiver that measures distance using the travel time of radio signals. There are several self-help guides (Letham 1998; Anderson 2002) if one desires a more in depth understanding.

To acquire, display, compile, and interpret data that has been recorded in the field, a geographic information system (GIS) is commonly used. GIS technology allows for multiple projections with some ease in converting from one to another, and records captured by GPS can quickly be georeferenced. Multiple spatial data layers such as administrative boundaries for state(s) and counties can reside within a GIS; if deemed important, they can be used as checks in the accuracy and data quality process. Spatial technologies provide tools to incorporate and analyze large data sets in a meaningful manner with the production of useful information. Data can be converted or displayed by locations or across a landscape and displayed as charts, drawings, or maps. These technologies provide a means to handle complexities such as incorporating scale and hierarchy concepts into ecosystem-based management approaches (O'Neill et al. 1996). These technologies also allow others to see how decisions are made, thus leaving a footprint(s) in the decision making process to follow.

Spatial technologies and mapping are as important to the manager as calculators and vehicles. Appendices 2–6 give examples of the spatial components for location and time data elements associated with differing levels of project complexity. Thus, spatial technologies can provide timely information in usable formats for decision makers. Spatial technologies like GIS are frequently described in terms of hardware (computers and workstations) and software (computer programs). Typically more computing power (speed and memory) and large data storage (disk space) are preferred. Workstations do most of the heavy lifting in handling large and/or complex data sets; to write and transfer data effectively in and out of these systems requires peripherals like tape storage and retrieval systems, CD-ROM, and DVD-RAM writers. To become familiar with developing applications using these technologies, see O'Neil et al. (2005).

Emerging Global Data Systems

In the following paragraphs, we provide an overview of some of the key global data systems relevant to salmonid researchers and managers.

The Ocean Biogeographic Information System (OBIS) is a Web-based provider of globally georeferenced information, a project of the Census on Marine Life, which plans to explain global biological diversity over 10 years with the collaborative efforts of 1,000 researchers in 73 nations. GBIF plans to be the "most-used gateway to biodiversity and other biological data on the Internet" by 2011. OBIS requires only that each data set contain the following four common fields: latitude, longitude, taxonomic name, and date/time of last modification; GBIF's requirements are similar.

State of the Salmon, a joint project of Wild Salmon Center and Ecotrust, is undertaking a salmon monitoring data inventory throughout the North Pacific; it focuses on large polygons such as Salmon Ecoregions and Hydro 1K (<www.stateofthesalmon.org/pattern/page.asp?pID=62>); because information is digitized, smaller data sets can easily be rolled into these larger regions. For example, State of the Salmon's database of salmon distribution at the Pacific scale incorporated data from the Raincoast Conservation Society, which works along the central coast of British Columbia at a very fine scale. Raincoast surveyed streams at specific field stations to record presence/absence of several species; since Raincoasts's data were collected with latitude/longitude coordinates, State of the Salmon easily overlaid the point coverage on top of its distribution coverage to make corrections to its database.

Through its volunteer membership of 7,000 species conservation experts, the Species Survival Commission (SSC) of the World Conservation Union (IUCN) holds what is probably the world's most complete body of information on the status and distribution of species threatened with extinction. Although abundant, the data and information contained within the SSC network is widely dispersed and sometimes difficult to access. More than half of SSC's members reside in the developing world and experience constraints in ability to store, share, and analyze their data. The Species Information Service (SIS) is being developed as SSC's data management initiative to address this problem. The SIS aims to become a worldwide species information resource (interlinked databases of species-related information managed by SSC's network of specialist groups). It will be easily accessible to the conservation and development communities, including scientists, natural resource managers, educators, decision makers, and donors, and will contribute to integrated biodiversity conservation products.

The core business of the World Conservation Union (IUCN) is generating, integrating, managing, and disseminating knowledge for conservation. This knowledge is used to empower people and institutions to plan, manage, conserve, and use nature and natural resources in a sustainable and equitable manner. IUCN is undertaking a new communications initiative: the Green Web. This will be IUCN's portal to all of its information resources, with SIS as the backbone. The SIS enables and empowers IUCN expert networks to bridge the scientific digital divide. In turn these networks empower educational institutions, nongovernmental orgainzations, and local communities to make use of their information. In addition, through the Green Web, IUCN's regional and country offices and specialist groups will provide facilities for under-resourced groups and communities to access the Internet and IUCN's knowledge base.

Despite the significant advances in local, regional, and global data systems, disparities exist among countries and among groups within countries regarding access to and use of information and communications technology. This results in a digital divide among scientists and other experts working to conserve the world's biodiversity. The establishment, coordination, and long-term support for a network of regional data centers will significantly advance our collective conservation efforts.

Literature Cited

Anderson, B. 2002. GPS afloat--gps navigation made simple. Fernhurst Books, West Sussex, UK.

Duffy, W. 2003. California Coastal Salmonid Restoration Monitoring and Evaluation Program—interim restoration effectiveness and validation monitoring protocols. B. W. Collins, editor. California Department of Fish and Game, Fortuna, California.

Harding, S. M., and G. G. Wilkinson. 1996. A strategic view of GIS research and technology development for Europe. Space Applications Institute--Environmental Mapping and Modeling Unit, report of the expert panel convened at the joint research centre, Ispra, Italy.

Letham, L. 1998. GPS made easy. Second edition. The Mountaineers, Seattle.

NWQMC (National Water Quality Monitoring Council). 2004. Revised guide for data elements for reporting water quality monitoring results for chemical, biological, toxicological, and microbiological analytes. Methods and Data Comparability Board of the National Water Quality Monitoring Council, NWQMC Technical Report.

O'Neill, R. V. 1996. Recent developments in ecological theory: hierarchy and scale. Gap analysis – a landscape approach to biodiversity planning. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland.

O'Neil T. A., P. Bettinger, B. G. Marcot, B. W. Luscombe, G. T. Koeln, H. J. Bruner, C. Barrett, J. A. Pollock, and S. Bernatus. 2005. Application of spatial technologies in wildlife biology. Pages 418–447 in C. Braun, editor. Wildlife techniques manual, 6th edition, Bethesda, Maryland.

Paulus, J., and S. Toshach. 2006. Best practices for recording location and time related data. White Paper–Northwest Environmental Data-Network. Northwest Power and Conservation Council, Portland, Oregon.

Prévost, Y. A., and P. Gilruth. 1997. Environmental information systems in sub-Saharan Africa. Towards environmentally sustainable development in sub-Saharan Africa—building blocks for Africa 2025. The World Bank, UNDP Post-UNCED Series Paper 12, Washington, D.C..

van Genderen, J. L. 1991. Guidelines for education and training in environmental information systems in sub-Saharan Africa: some key issues. The World Bank International Advisory Committee Guidelines Series 2, Washington, D.C.

Appendix A: Global data elements within the Darwin Core project

Element	Description	Can be NULL	Type	Min. Value	Max. Value
Record-level elements					
Global unique identifier	A universal resource name for the global unique identifier for the specimen or observation record	No	String		
Date last modified	Last time the data for the record was modified (e.g., June 5, 1994, 8:15 a.m., U.S. Eastern Standard Time)	No	Date-Time		
Institution code	Code or acronym identifying the institution administering the collection	No	String		
Collection code	Code or acronym identifying the collection within an institution in which the record is cataloged	No	String		
Catalog number	The alphanumeric value identifying an individual organism record within the collection	No	String		
Taxonomic Elements					
Scientific name	Full name to lowest taxon that the organism can be identified	No	String		
Kingdom	Name of kingdom where organism is classified	Yes	String		
Phylum	Name of the phylum (or division) of organism.	Yes	String		
Class	Name of the class of organism	Yes	String		
Order	Name of the order of organism	Yes	String		
Family	Name of family of organism	Yes	String		
Genus	Name of genus of organism	Yes	String		
Locality Elements					
Continent	Full name of the continent	Yes	String		
Water body	Full name of the body of water	Yes	String		
Island	Full name of the island	Yes	String		
Country	Full name of the country	Yes	String		
State/province	Full name of the state, province, or region	Yes	String		
Locality	Description of the locality where collection occurred	Yes	String		
Minimum elevation in meters	Minimum altitude above (positive) or below (negative) sea level	Yes	Double		
Maximum elevation in meters	Maximum altitude above (positive) or below (negative) sea level	Yes	Double		
Minimum depth in meters	Minimum depth below the surface of the water	Yes	Double		

Element	Description	Can be NULL	Type	Min. Value	Max. Value
Maximum depth in meters	Maximum depth below the surface of the water	Yes	Double		
Geospatial Elements					
Decimal latitude	Latitude of the collection location shown in decimal degrees.	No or could use other locational information	Double	-90	90
Decimal longitude	Longitude of the collection location shown in decimal degrees.	No or could use other locational information	Double	-180	180
Geodetic datum	Geodetic datum the latitude and longitude refer	Yes	String		
Collecting Event Elements					
Year collected	4-digit year in the Common Era calendar	Yes	Gregorian Year		
Month collected	2-digit month of year in the Common Era calendar	Yes	Integer	1	12
Day collected	2-digit day of the month in the Common Era calendar	Yes	Integer	1	31
Time collected	Time of day of collection or observation	Yes	Double	0	< 24
Julian day	Ordinal day of the year	Yes	Integer	1	366
Collector	Name(s) of collector(s)	Yes	String		
Biological Elements					
Sex	The sex of a biological individual	Yes	String		
Life stage	The age class, reproductive stage, or life stage of the organism	Yes	String		
References Elements					
Image URL	Digital images associated with the specimen or observation	Yes	String		
Related information	References to information	Yes	String		

Appendix B: Spatial Components for General Location and Time Data Elements Associated with Project—Level 1: General Project Information

Logical name	Name definition	Element name (for example only)	Element code, code range, or description
Project	A project is a unit of work defined by an organization or entity. A project may include one or more sites or one or more types and number of activities. Unique system identifier	PRJ_ID	Examples • Skagit River Habitat Restoration Project • Okanogan Water Quality Sampling Project • Oregon North Coast Nearshore Monitoring Project • Deschutes River Flow Monitoring Project
Project location description	Term that best describes the field location in relation to the surrounding environment	PRJ_LOC_DESC	Text field Examples: • Okanogan watershed • ESA Region • SW 1/4 of Section 36 of Township 29 Range 01
Project location latitude coordinate	Distance north or south of the equator. Decimal equivalent to the degrees-minutes-seconds latitude value	PRJ_LOC_LAT_COORD	Float, 2 places, 6 decimals; (4 decimals minimum) E.g. Range for WA: 45.000000 – 49.999999
Project location longitude coordinate	Distance east or west of the Central Meridian (Greenwich, England). Decimal equivalent to the degrees-minutes-seconds longitude value	PRJ_LOC_LONG_COORD	Float, 3 places, 6 Decimals, will accommodate signed values (4 decimals minimum); E.g. Range for WA: -116.000000 – -125.999999
Project horizontal datum	Model used to match the horizontal position of features on the ground to coordinates and locations on a map. (Note: When taking GPS measurements, it is very important to record your datum!)	PRJ_HORZ_DAT	01 - N. American Datum 1927 (NAD27- used on many USGS quad maps or NOAA charts); 02 - N. American Datum 1983 (NAD83 or 91 Adj.—based on Earth and satellite observations, similar to WGS84 but specific to North America.); 03 - High Accuracy Reference Network (HARN—similar to NAD83, but more accurate per GPS observations); 04 - World Geodetic System of 1984 (WGS84—world datum, based on Earth and satellite observations); 99 - unknown.
Project location collection method	Technique used to collect the horizontal coordinates of a Location	PRJ_LOC_COLL_MTH	1 - Address Matching - Block Face; 2 - Address Matching - House Number; 3 - Address Matching - Street Centerline; 4 - Address Matching - Unknown; 5 - Aerial Photography - Rectified; 6 - Aerial Photography - Unknown; 7 - Aerial Photography - Unrectified; 8 - Cadastral Survey (conventional land survey); 9 - Census Block 1990 Centroid; 10 - Census Block Group 1990 Centroid; 11 - Conversion from STR; 12 - Digital or manual raw photo extraction; 13 - Digitized off CTR screen/digitial data; 14 - Digitized - paper map; 15 - GPS carrier phase (employs the satellite Code's carrier signal to improve accuracy); 16 - GPS code phase (measurements based on pseudo random code broadcast by satellite); 17 - GPS kinematic (tracking location while moving using carrier phase); 18 - GPS (Unknown); 19 - Hand-measured - paper map (interpolation); 20 - LORAN-C; 21 - Orthophotography - digital; 22 - Orthophotography - paper; 23 - Satellite Imagery - Landsat MSS (Multi-Spectral Scanning); 24 - Satellite Imagery - Landsat TM (Thematic Mapper); 25 - Satellite Imagery - Other; 26 - Satellite Imagery - SPOT Panchromatic; 27 - Satellite Imagery - SPOT Multi Spectral; 28 - Zip Code Centroid; 29 - GPS (Code/Differential); 30 - Estimated Value 99 - unknown

Logical name	Name definition	Element name (for example only)	Element code, code range, or description
Project start date	The date that the project activity commenced		Date, YYYY/MM/DD format. (Only if applicable) E.g. 2003/03/12. Use a date of 1800/01/01 to indicate that the Start Date is not specified or is unknown.
Project end date	The date that the project activity ended		Date, YYYY/MM/DD format. (Only if applicable) E.g. 2004/03/12. Use a date of 1800/01/01 to indicate that the End Date is not specified or is unknown.

(This detail may not be necessary for all reporting purposes)

Note the information presented here is from Paulus and Toshach (2006).

Appendix C: Spatial components for location and time data elements associated with project—Level 2: project tracking at specific or numerous sites over time

Logical name	Name definition	Element name (for example only)	Element code, code range, or description
Project	The place where site activities associated with a project occur or the area where the work is done. Each site will pertain to just one project, but there can be more than one site for any given project. Location of the site or activities where work is conducted (on-the-ground activities) Unique system identifier	PRJ_SITE_ID	Note to readers: These elements need to be defined based on the type of project site work that is being done • Skagit River Habitat Restoration Site—2 stream reaches • Okanogan Water Quality Sampling Site—4 monitoring sites in study • Oregon North Coast Nearshore Monitoring Site—3 coastal reaches in project • Deschutes Flow Monitoring Site—2 gauging stations in project
Project location description	Term that best describes the site location in relation to the surrounding environment. Information that describes the place a Location exists	PRJ_SITE_LOC_DESC	Text field Example: • 200 yards north of the cattle crossing on Laumann Road, north of the intersection with Heidi Road
Project location latitude coordinate	Distance north or south of the equator. Decimal equivalent to the degrees-minutes-seconds latitude value	PRJ_SITE_LOC_LAT_COORD	Float, 2 places, 6 decimals; (4 decimals minimum) Eg., Range for WA: 45.000000-49.999999
Project location longitude loordinate	Distance east or west of the Central Meridian (Greenwich, England). Decimal equivalent to the degrees-minutes-seconds longitude value	PRJ_SITE_LOC_LONG_COORD	Float, 3 places, 6 decimals, (4 Decimals minimum); will accommodate signed values; Eg., Range for WA: -116.000000 – -125.999999
Project horizontal datum	Model used to match the horizontal position of features on the ground to coordinates and locations on a map (Note: When taking GPS measurements, it is very important to record your datum!)	PRJ_SITE_HORZ_DAT	01 - N. American Datum 1927 (NAD27- used on many USGS quad maps or NOAA charts) 02 - N. American Datum 1983 (NAD83 or 91 Adj.—based on Earth and satellite observations, similar to WGS84 but specific to North America) 03 - High Accuracy Reference Network (HARN—similar to NAD83, but more accurate per GPS observations) 04 - World Geodetic System of 1984 (WGS84—world datum, based on Earth and satellite observations) 99 - unknown
Project location collection method	Technique used to collect the horizontal coordinates of a site location	LOC_COLL_MTH	See Appendix 2 for potential list
Project start date	The date that the project activity commenced		Date, YYYY/MM/DD format. (Only if applicable) E.g. 2003/03/12. Use a date of 1/1/1800 to indicate that the Start Date is not specified or is unknown.
Project end date	The date that the project activity ended		Date, YYYY/MM/DD format. (Only if applicable) E.g. 2004/03/12. Use a date of 1800/1/1 to indicate that the End Date is not specified or is unknown.

(This detail may not be necessary for all reporting purposes)

Note the information presented here is from Paulus and Toshach (2006).

Appendix D: Spatial components for location and time data elements associated with project—Level 3: complex projects that track specifically measured features at a site

Logical name	Name definition	Element name (for example only)	Element Code, Code Range, or Description
Site Feature	The structure, form, or appearance of what is being tracked, measured or observed at any given project site. Within any give project site there may be various features represented as single points, linear features or aerial extents. Unique system identifier	SITE_FEA_ID	Note to readers: This needs to be defined based on the type of scientific/field information that is being collected *Example Code Tables:* • Transect measurement point • Fence • Wells • Fish hatchery raceway • Reach segments *Examples of Site Features:* Water sampling well locations • Individual gauging station location Location of addition to spawning gravel
Site feature location description	Term that best describes the feature location in relation to the surrounding environment. Information that describes the place a Location exists	SITE_FEA_LOC_DESC	Text field Example: 200 yards north of the cattle crossing on Laumann Road, north of the intersection with Heidi Road
Site feature location latitude coordinate	Distance north or south of the equator. Decimal equivalent to the degrees-minutes-seconds latitude value	SITE_FEA_LOC_LAT_ COORD	Float, 2 places, 6 decimals; (4 decimals minimum) Eg. Range for WA: 45.000000 – 49.999999
Site feature location longtitude coordinate	Distance east or west of the Central Meridian (Greenwich, England). Decimal equivalent to the degrees-minutes-seconds longitude value	SITE_FEA_LOC_LONG_ COORD	Float, 3 places, 6 decimals, (4 decimals minimum); will accommodate signed values; e.g., Range for WA: -116.000000 – -125.999999
Logical site feature horizontal datum	Model used to match the horizontal position of features on the ground to coordinates on a map	SITE_FEA_HORZ_DAT	01 - N. American Datum 1927 (NAD27- used on many USGS quad maps or NOAA charts) 02 - N. American Datum 1983 (NAD83 or 91 Adj.—based on Earth and satellite observations, similar to WGS84 but specific to North America.) 03 - High Accuracy Reference Network (HARN—similar to NAD83, but more accurate per GPS observations) 04 - World Geodetic System of 1984 (WGS84—world datum, based on Earth and satellite observations) 99 - unknown
Site feature start date	Technique used to collect the horizontal coordinates of a feature location.	SITE_FEA_STR_DT	Date, YYYY/MM/DD format. (Only if applicable) E.g. 2003/03/12. Use a date of 1800/1/1 to indicate that the Start Date is not specified or is unknown
Site feature end date	The date that the feature activity (sample collection, field measurement, field observation) ended. If a feature activity is essentially instantaneous, a Feature End Date is often not specified.	SITE_FEA_END_DT	Date, YYYY/MM/DD format. (Only if applicable) E.g. 2004/03/12. Use a date of 1800/1/1 to indicate that the End Date is not specified or is unknown
Site feature start time	The time that the feature activity began, for example the time of sampling	SITE_FEA_ST_TM	Feature start time (time the collection, measurement, observation started -using a 24hr clock at local time) (hhmmss) e.g., 164322 (only if applicable)
Site feature end time	The time that the feature activity ended, for example the end of sampling	SITE_FEA_END_TM	Feature end time (time the collection, measurement, observation ended -using a 24hr clock at local time) (hhmmss) e.g., 175231 (only if applicable)

(This detail may not be necessary for all reporting purposes)

Note the information presented here is from Paulus and Toshach (2006).

Appendix E: Optional elevation data associated with projects, sites, or a feature

Logical name	Name definition	Element name (for example only)	Element code, code range, or description
Elevation	The measure of the elevation of the project site above a reference datum	PRJ_SITE_VERT	Float, will accommodate signed values
Elevation units	The unit of measurement used to describe the elevation value	PRJ_SITE_VERT_UNIT	Text field; example Meters Feet
Elevation datum	The code for the reference datum used to determine the vertical measure	PRJ_SITE_VERT_DAT	Navd88 Ngvd29 Mean Sea-Level Local tidal datum Other
Elevation collection method	The technique used to establish the elevation or depth of the sampling site	PRJ_SITE_VERT_COLL_MTH	GPS carrier phase static relative position GPS carrier phase kinematic relative position GPS code (pseudo range) differential GPS code (pseudo range) precise position GPS code (pseudo range) standard position (Sa off) GPS code (pseudo range) standard position (Sa on) Other Altimetry Precise leveling-bench mark Leveling-non bench mark control points Trigonometric leveling Photogrammetric Topographic map interpolation

Note the information presented here is from Paulus and Toshach (2006).

Appendix F: Examples of location and time reporting for different types of features

Feature Name	Examples of location/time reporting detail from independent data collectors	Examples of location/time reporting detail from corporate data collectors
Install fish screen	• Location of screen (lat./long. dec. degree) • Date of install: YYYY/MM/DD	• Location of screen (lat./long. dec. degree) • Date of install: YYYY/MM/DD
Stream bank stabilization	• Start and end point of stabilization (lat./long. dec. degree) • Date of stabilization: YYYY/MM/DD	• Polygon of stabilization area (lat./long. dec. degree) • Date of stabilization: YYYY/MM/DD
Riparian area treated	• Start and end point (lat./long. dec. degree) • Date of treatment: YYYY/MM/DD	• Polygon of area treated (lat./long. dec. degree) • Date of treatment: YYYY/MM/DD
Road obliteration project	• Start and end point (lat./long. dec. degree) • Length of treatment • Date of obliteration: YYYY/MM/DD	• Line detail of road treatment (lat./long. dec. degree) • Date of obliteration: YYYY/MM/DD
Sediment control basin	• Centroid of basin (lat./long. dec. degree) • Date of sediment control: YYYY/MM/DD	• Polygon of basin (lat./long. dec. degree) • Date of sediment control: YYYY/MM/DD
Wetland creation project	• Centroid (lat./long. dec. degree) • Date of wetland creation: YYYY/MM/DD	• Polygon (lat./long. dec. degree) • Date of wetland creation: YYYY/MM/DD
Invasive species treatment	• Centroid of treatment area • Date of treatment: YYYY/MM/DD	• Polygon of treatment area (lat./long. dec. degree) • Date of treatment: YYYY/MM/DD
Hatchery fry/smolt release	• Location of point of release (lat./long. dec. degree) • Date and Time of release YYYY/MM/DD, hhhh/mm/ss	• Location of point of release (lat./long. dec. degree) • Date and Time of release YYYY/MM/DD, hhhh/mm/ss
Sampling site	• Location (lat./long. dec. degree) • Date and time of sample: YYYY/MM/DD, hhhh/mm/ss	• Location (lat./long. dec. degree) • Date and time of sample: YYYY/MM/DD, hhhh/mm/ss
Livestock exclusion fencing	• Start and end point (lat./long. dec. degree)	• Line detail (lat./long. dec. degree)

Note the information presented here is from Paulus and Toshach (2006).

Methods

This handbook was inspired by a vision to provide standard methods for the capture and counting of salmonids and to serve as the foundation for consistent regional and global data sets describing salmon populations.

The handbook aims to establish standard methods for four sampling objectives: (1) abundance, (2) distribution, (3) population trends, and (4) fish/habitat relationships. The initial step for this resource entailed a literature review of published and unpublished protocols for all commonly used sampling methods in freshwater habitats. The literature search focused on state and federal agencies, universities, and tribes within the Pacific Northwest and elsewhere in North America. Experts were identified and asked to provide available protocols for fish capture and counting. At universities, we approached faculty and research staff directly and conducted searches of university library resources.

We searched existing databases with protocols such as those maintained in British Columbia and the Klamath Resource Information System. Reference sections from available protocols were also used to track down original documents. After assembling more than 375 documents, we used a specific set of criteria to conduct a coarse screening on the content and value of each document. The screening process, based on work by Oakley et al. (2003), used the checklist below to determine whether protocols met a minimum standard for further consideration.

Essential Elements of Protocols

Background and objectives
Background—history, resources being addressed
Rationale—justification of selecting a given resource to inventory or monitor
Objective—list of measurable tasks
Sampling design
Site selection—defining boundaries or "populations" sampled; selecting sampling locations; stratification; and spatial design
Sampling frequency and replication—recommended number and location of sampling sites; frequency and timing of sampling; level of change that can be detected for the amount/type of sampling
Field/office methods
Setup—field season preparations and equipment setup (including permitting/compliance procedures)
Events sequence—sequence of events during field season or during preparation of a monitoring plan
Measurement detail—details of taking measurements, with examples of field forms
Sample processing—postcollection processing of samples (e.g., lab analysis, preparing voucher specimens)
Data handling, analysis, and reporting
Metadata procedure—database fields and sizes; sample collection information; site description; quality assurance
Database design—overview of database design and structure illustrating relationships between tables
Data entry—data entry procedures; verification and editing of data
Data summaries—data summaries and procedures for conducting statistical analyses
Report format—recommended report format with examples of summary tables and figures
Trend analysis—recommended methods for trend analysis
Archival procedures—data archival procedures
Personnel requirements and training

Background and objectives
Responsibilities—crew and project roles and responsibilities
Qualifications—prior experience for paid and volunteer staff
Training—availability, locations, timing, and procedures
Operational requirements
Workload and schedule—factors to facilitate chronological planning
Equipment needs—list of equipment, materials, and facilities needed
Budget considerations—calculation guidelines
Literature cited

The coarse review occurred over several months, with several fish biologists reviewing documents. Each protocol described in a document was evaluated against the above criteria and was rated as "fully covered," partially covered," or "not covered at all." The criteria and ratings for each protocol were stored in digital data sheets and organized in a searchable database along with author and citation information. The ratings revealed desirable components of each protocol and determined which were recommended for standardization.

From the coarse review, more than 375 documents describing protocols were reduced to 74 of the strongest evaluated. These "strongest 74" reflected the highest ratings for essential elements of protocols and were distributed among the different methods identified as those commonly used for capture and counting of salmonids and other freshwater fish. Through this process, the reviewers affirmed that for most methods no single document contained information meeting all criteria. Thus, to get protocols with the full range of elements, the strongest parts of existing documents would need to be highlighted and combined. If none of the documents reviewed were found to have essential elements, the authors decided to develop them from scratch.

Following literature review and protocol screening, a number of international experts in fish capture and population monitoring were invited to attend a workshop in Welches, Oregon, in March 2004. Participants were selected based on their professional and geographic areas of expertise from around the Pacific Rim and grouped into three work groups—field practitioners, biometricians, and data managers. The groups received a set of 375 protocols identified from the literature review and were asked to identify the most promising and pertinent sections. Each group drafted protocols relevant to its area of specialization. For example, the field practitioners drafted protocols for several data collection methods, the biometricians focused on data analysis methods, and data managers prepared a core set of variables and units for those variables critical to any fish population monitoring effort. The workshop contributed immensely to validating and refining best practices for an array of data collection, analysis, and management methods. The panelists also highlighted methods (e.g., counting towers, cast netting) not reflected in the initial literature search that were subsequently added to the final set of protocols and reviewed.

Workshop results were turned over to the core project team for further editing, peer review, and writing. A substantial task then confronted the project team: many of the protocols needed significant development and refinement. Teams of authors, consisting of individuals recognized as experts in their respective methods, including some workshop participants, were enlisted to further develop and refine the protocols. This process produced a set of draft 13 primary protocols, 5 supplemental techniques, and several essays, all of which were sent out for formal peer review.

Literature Cited

Oakley, K. L., L. P. Thomas, and S. G. Fancy. 2003. Guidelines for long-term monitoring protocols. Wildlife Society Bulletin 31:1000–1003.