Identifying Best Available Science*

Richard B. Shepard, Ph.D.†

Executive Summary

The requirement to use the best available science appears in most environmental regulations at the federal and state levels, but best available science is not defined in the dictionary or by consensus. Washington state lists excellent criteria in its Administrative Code (WAC 365-195-905), but these criteria are not complete nor easily applied by the non-technical decision-maker.

This white paper explains why there is no simple definition of best available science then offers a hierarchy of criteria that can be applied by decision-makers to satisfy both regulatory requirements and public concern. Thoughtfully applied, these criteria could also resolve the perennial conundrum of dueling scientists. There are four reasons why best available science cannot be neatly packaged into a precise definition. These reasons are: 1) the dynamic nature of natural ecosystems, 2) importance differences among terrestrial, aquatic, riparian, arid, semi-arid, humid, and other ecosystem types each having a different best, 3) our inability to completely characterize an ecosystem's structure and functions, and 4) the subjective, value-based definitions of best. Despite these limitations valid political and administrative decisions can be made that fulfill the need of applying best available science. While there will almost always be one or more stakeholder groups who do not like the outcome, a fully-documented decision would be technically sound and legally defensible.

The approach proposed in this paper involves three categories of data: quantity, quality, and relevance to the decision to be made. Criteria are presented as a set of seven questions so that non-scientists can apply them with confidence. While the assistance of an objective scientist can be very helpful, it is not necessary if the decision-makers are careful and completely document their evaluation.

1 Introduction

Virtually every federal and state statute and regulation requires that land use, industrial project, and development decisions be based on the best available

†President, Applied Ecosystem Services, Inc.; Troutdale, OR U.S.A.

science. That the definition of this term remains elusive is a reflection of the complex, dynamic nature of natural systems and the reality that they are all different. What is considered the best available science for fill or removal in a jurisdictional wetland is not the best available science for understanding salmon population dynamics in a river system, or for mining or energy development projects at upland sites. Differences among ecosystems is one reason why a consensus definition of best available science cannot be found. This is the most important point: best available science is always within a defined context of the decision to be made.

There are three categories that must be evaluated when establishing the best available science applicable to a specific project at a specific location and at a specific time. These categories are data quantity, data quality and data relevance. Washington state's *Best Available Science Rule* (WAC 365-195-900–365-195-925) touches on these three categories in a different way, but the Rule assumes objectivity and quality that are based more on conventional wisdom than objective data. These deficiencies can be corrected by applying alternative criteria. These alternative criteria are presented in this white paper as questions that decision-makers should answer.

It is important to keep in mind the relationships that underlie the desire to identify the best available science. Data by themselves are useless; they are only the bricks and not the entire structure. Appropriate analytical techniques transform the raw data into information; the walls of the structure. When the information is interpreted for meaning in a specific situation it becomes knowledge, and is then the basis for decision making: the completed and furnished structure.

The purpose of this white paper is to stimulate thought and discussion on how to identify the best available science germane to a permit application, land use determination, or policy decision. It is intended for use at all levels of government by regulators, administrators, and policy makers.

2 Data Quantity

The category of data quantity is more than the volume of data presented in support of one position or another. However, the amount is pertinent to the inferences that can be drawn from the existing data. In many situations the only available data are those collected in support of a permit application or policy decision (e.g., inventories conducted under programs such as Oregon's Statewide Planning Goal 5 or baseline studies related to the federal NEPA process). If these are the only available data they do not automatically become the best available science. The pedigree of the data need to be examined before a decision can be made.

The size of the available data sets are one consideration. Was the pertinent area well-covered by sampling sites? Were samples collected in one season or over a longer time span? Were the methods and equipment used of adequate resolution to provide suitable data? What biases or selective pressures are associated with the methods used? If data were collected over long time periods, are they comparable? These are a solid start to many questions that can be asked about data quantity and used to support a decision on the best available science.

2.1 Areal Coverage

Field science students are taught the rudiments of sampling design. Among the statistical designs frequently taught are random, stratified, stratified-random and Latin square. Each of these approaches provide for spatial placement that produce data that can be statistically analyzed. However, they frequently do not work in the real world as well as they do in theory.

In terrestrial environments most cover types (i.e., vegetation) are not randomly distributed across a site. Trees may be restricted to draws in hillsides; south-facing slopes are very different from north-facing slopes, riparian vegetation in semi-arid climates may be absent or very different from that beyond the fringe. Forest edges in humid environments offer vegetational variations in abundance compared with the interior (particularly in closed-canopy, mature stands). To sample vegetation, soils, mammals, invertebrates or whatever is of interest from the site may require the deliberate placement of sampling locations to capture the range of environments present. This placement may well not fit into any statistical sampling design. However, properly analyzed and interpreted the data are even more meaningful than that collected by following a cookbook approach. Terrestrial environments tend to be relatively stable places with change occurring slowly over long times.

Aquatic environments differ in all aspects from terrestrial ones. They are highly dynamic; streams and rivers even more than lakes and reservoirs. Most data collection in water larger than headwater streams is done blind: we cannot see the bottom or the water much below the surface. Aquatic animals are exceptionally good at avoiding capture regardless of whether it is a fish (just ask any fisherman!) or insect. One of the best known examples of the latter is the large stonefly nymph with the common name of Giant salmonfly *(Pteronarcys californica*). This large insect (mature stages can be 5 cm [2 inches] in length) can dig down in streambed gravels faster than a collector can dig after them.

Another problem with aquatic sampling is that all equipment is biased toward some size animals and against other size animals. This is most easily seen in nets for fish, invertebrates and plankton. A large mesh size lets smaller individuals swim through the holes and not be part of the collection. A small mesh traps the little individuals but can create a back wash of water that pushes out some individuals, particularly the larger and more mobile ones.

A third problem with aquatic sampling is an extension of the above noted conditions. Specifically, in aquatic environments we can gather a *collection* but not necessarily a statistically-valid *sample*. The difference is important during data analyses. Many statistical tests are based on the assumption that the data points are normally distributed. That is, if plotted and connected by the best-fit line, the points would form the familiar bell-shaped curve. Also, extrapolation from the sample to the entire population is predicated on the sample being truly representative of the entire population. However, with a collection of individuals the assumption that they represent the proportion of species in the entire population cannot be met. This means that the class of statistical procedures called "non-parametric" are the appropriate ones to use. There are non-parametric equivalents for almost all the parametric statistical tests that anyone would want to use on field-collected data.

Fish and wildlife data often come from creel surveys, commercial catch reports, tag returns and other indicators of the number of animals removed from the population. These data are insufficient to support any conclusions about overall population size unless there have been concurrent studies that document the percentage of the population that was captured and removed. Methods of data collection are, therefore, important in deciding whether or not sufficient data exist to be considered the best available science.

Question 1: Am I satisfied that sufficient samples were collected to adequately cover the entire area under consideration?

2.2 Temporal Coverage

When were the data collected? And, for how long? If plants need to be identified and counted the most complete data are collected during the growing season. Most aquatic insects in streams and rivers are present as juveniles of sufficient size to collect and identify during the winter. Birds are much easier to find and identify by sight and sound during their mating season. Water chemistry samples should allow measurement of variation in three dimensions. While it is rare to have the luxury of collecting field data only during optimal times, the potential effects on the quantity of data need to be considered when the data are used to support a decision. Knowing when data are collected relative to what those data are makes it easier to evaluate if there is a sufficient quantity to support conclusions.

When the questions being asked are about change over time, then the appropriate data reduction method is that of time-series analyses. These analyses can be conducted to remove seasonal effects and reveal overall trends. However, one of the requirements of this analytical method is that the data must be collected at constant time intervals. It is not permissible to collect data monthly for the first six months of the study, quarterly for the rest of that first year, then semi-annually for the next two years. While this does yield data over a three-year period, the unequal sampling intervals precludes the use of time-series analyses.

It is common to find data from public agencies, university research, or anecdotal origins that were collected incidental to other projects. These might be used to expand the temporal coverage of the project of interest if adequate justification for their use is provided. Such justification includes an assessment of the circumstances under which the data were collected. For example, the powerhouse and fish bypass facility operated by Portland General Electric at

Willamette Falls (Willamette River, Oregon) has an operator recording fish passage counts, but only when the turbines are in operation. Therefore, these data may represent only a portion of a day, at infrequent or patchy intervals and not part of a program designed specifically to evaluate fish passage past the dam. These data certainly can be used to augment other fish counts, but they are insufficient for some uses, such as estimating anadromous salmonid returns to the river above the falls or the number of migrating juveniles that pass this point along the river.

Question 2: Am I comfortable that the data cover enough time to understand variation with time?

2.3 Methods

Technology and experience change how data are collected. Birds used to be collected by shooting with shotguns loaded with very small pellets ("museum dust"), fish population estimates based on the size of the commercial harvest, and water chemistry measured with wet laboratory procedures carried out in the field or on preserved (when necessary) samples returned to the laboratory. Now bird populations can be counted by sound and sight with representatives "captured" as digital images. Wildlife trigger camera/flash units set up along game trails, river reaches are seined, electrofished or rotenoned to count individuals and water chemistry is measured using analog or digital meters that permit the measurement of values at different depths in the water column in real time.

A concern with incorporating historic data with current data is adjusting for biases or efficiencies of different methods. This can most often be done, but it is not sufficient to report numbers separated by decades as if they are equivalent. More significant are the differences produced when manual methods are compared with automated methods.

Toward the end of the last decade the Nevada Division of Environmental Protection's Water Quality Bureau determined that summer water temperatures in the Humboldt River exceeded state and federal standards in the eastern portion of the state (in the Carlin-Elko area). During discussions with staff it was revealed that data were collected by having a technician measure water temperature in the river at the sinks east of Reno early in the morning and continue taking samples at specific locations along the 350-mile drive toward the headwaters. Almost invariably, it was mid- to late-afternoon before the technician arrived at the Carlin and Elko sampling stations; the time of day when water temperature was near its maximum. Water quality decisions were made based on early morning samples to the west and late afternoon samples to the east. This time spam meant that there was no usable data on diel (i.e., daily) temperature ranges or rates of change at any sampling location in the river. When it was suggested that the Bureau use submerged, recording temperature data loggers anchored to the river bank (and well protected from vandalism) they changed to this method. Now they can send out a technician every month or two to retrieve each data logger, download the stored temperature data (measured at 0.5-hour intervals) onto a portable computer and have data collected at the same time at each location along a 350-mile river course. These data can now be considered the best available science for water temperature in the Humboldt River system.

Creel surveys of fish caught by anglers cannot be directly compared with in-stream surveys that count all—or most—of the existing population. The creel survey results reflect the skill of the angler and the intelligence of the fish, neither indicative of the entire fish population size. For mammals, birds and other terrestrial fauna, there are many methods available to census populations. It is reasonable to expect justification for combining data collected by different methods by having a section in the report that discusses each method in sufficient detail for the non-specialist to understand what portion of the entire population is represented by the numbers presented.

Question 3: Are the methods explained well enough that I can understand their limitations and value?

3 Data Quality

While the sequence: data... information... insight requires a solid base in terms of data quantity, quality, and relevance quality is sometimes the most difficult component to evaluate. Data quality refers to the way it was collected, analyzed, and interpreted. While this may seem to be a daunting task for the non-specialist in the specific scientific subject under consideration, it can be made accessible to everyone from the proper perspective.

The proper perspective is from the insight stage. What is the question being answered? Is this a resource allocation decision? Is the decision evaluating whether a specific activity might have unacceptable impacts on a natural resource? Is the goal to establish the population size of a plant or animal species under the Endangered Species Act? Once the purpose is clearly stated it is possible to ask what needs to be known in order to make an informed decision. If this is a question you cannot answer for yourself, then it is time to ask the technical domain expert. The two questions asked of the technical domain expert should be: 1) What information do I need to make a decision? and 2) How does this information relate to decision making? The same questions are asked about the information: what analytical methods will yield this information? and why are these the appropriate methods to use? After these questions are answered it is possible to evaluate the data to determine if they were collected using the appropriate techniques and temporal-spatial considerations that meet the assumptions of the analytic methods used.

Most of the valid scientific process characteristics listed in Washington's rule are quality issues (five out of six; the other falls in the category of relevance). The first characteristic in their list, peer review, is not an indicator of anything pertinent to the determination of best science. There is as much politics and going along with the herd in scientific publishing as there is in the corporate or government worlds. In every scientific subject there are numerous examples of poor, ethically questionable, and fraudulent research articles. The current one in the news is from England where the principal investigator of a study looking at the role of multiple disease inoculations (specifically, measles, mumps and rubella [MMR]) causing autism in children was partially funded by the attorney representing several families making that claim in court. The principal investigator's research results were published in the prestigious British medical journal, *The Lancet*. All of the principal author's coauthors have withdrawn their support of the conclusions. The lesson is to not trust a generalization ("peer reviewed journal publications always represent good science") and to make your own decision on a case-by-case basis.

Much more important than peer review with regard to identifying the best available science on natural ecosystem structure and function is a recognition of the dynamic nature of these systems. Examples will help clarify this importance. Most wetland boundaries are transitional zones that can be several meters wide. Placing wetland boundary markers anywhere within that zone is scientifically valid. Measuring water quality in a stream or river can produce different results from day to day, even from hour to hour. Two researchers can set up and run small mammal trap lines in the same field—and at the same location—on two separate nights and come up with different results. Both are correct because small mammals are not so consistent in their movements and use of trails that one can be assured of finding them on demand. It is vital that inherent variability be taken into account when evaluating any data or research. If the data source does not acknowledge this variability (directly or indirectly) then the data are suspect, peer reviewed or not.

Analytical methods tend to be standardized, but data collection methods are not. For example, in vegetation surveys in the semi-arid high desert environment of the Great Basin or Intermountain West, using a point sampler on a longitudinal transverse line may very well yield very low values for ground cover because a needle must intersect a plant part before that plant is counted. A better method might be to toss a 1-meter-per-side, plastic square randomly over one's shoulder and determine what percentage of the area inside the square is covered by vegetation. There are still more methods that may be used. Of greater value than a report on ground cover would be an assessment of what quantity and quality of habitats there are for various animal species. A statement that there is "abundant wildlife habitat" fails the analytical quality test. The data should be analyzed in terms of habitats for what types of animals and for what use (food, shelter, nesting, rearing young, etc.). Physical and chemical analyses are quite well defined under state and federal guidelines and standards. Somewhere near the data should be discussion of how the samples were collected, preserved (if needed), transported and analyzed. And, if a determination was made by a "windshield survey" (driving by the site) or stream channel condition assessed from the ridge tops rather than walking down each drainage, then the quality of the resulting data cannot be classified as best available. It may be the *only* available data, but it is neither best nor acceptable.

How the data are analyzed for significance is critically important in identifying best available science, but is difficult to evaluate for those without statistical training. There was one ground settling study a few years ago in the Portland, Oregon, metropolitan area in which the elevations of placed stakes were surveyed periodically. The one stake that increased in elevation was not included in the analysis because it was "obviously" wrong. The engineers who conducted the study then used multiple *t*-tests to analyze these data. This is a test of two sets of measurements to determine if the observed difference is due to chance and not the effect being investigated[1]. However, in the settling rate study, the question that should have been asked was: is the variation observed among all the stake measurements greater than the variations observed within each stake at different times? In other words, we want to look at the natural variation in measurements of a single stake taken at different times with the variation seen overall. The appropriate statistical test for this is called Analysis of Variance (ANOVA). The results of such a test would indicate whether the fill had completely settled. If you wanted to erect a building on this ground, the difference in results would be of great interest to you. Therefore, how the data are converted to information is very important in identifying the best available science. And, it needs to be evaluated for the data sources as well as any conclusions drawn from them.

Question 4: Are the right questions being asked, and the appropriate analytical and statistical methods applied to convert the data into information? If I cannot determine this for myself, I will find an expert who can, and who can explain everything to me so I truly understand it.

4 Data Relevance

Data relevance covers several considerations. Among these are whether the conclusions, inferences, and other transformations of information into knowledge are reasonable, understandable, and supported by similar studies at other times or other places. Are all supporting statements accompanied by references to their source? Perhaps of greatest importance, however, is the applicability of comparative studies to the project under consideration.

Any other study or data presented to support or refute a position needs to have its relevance completely and carefully documented. For example, if data from New Zealand, eastern Montana or any other location is cited, then it must be shown why it is applicable to the project location being considered. This context is a very important consideration in determining the best available science.

There will never be a time when we have all the data and insight possible about an ecosystem or the effects of any particular activity on the system or its components. This does not prevent informed decisions when those decisions are made with caution and careful explanation. There are unfortunate

[1]To clarify: you weight all the children in the fourth grade at an elementary school and you find that the boys' average weight is 68 pounds and the girls' average weight is 72 pounds. You want to know: is this difference significant or caused by chance? A *t*-test will tell you the answer.

occasions when data about a system are demanded before a decision is made whether to permit a certain activity, but it turns out that there is no established plan for the use of those data in the decision making process. It is reasonable to ask how data are to be used when they are requested.

An aspect of data relevance that is not considered in every situation is the effect of societal values and belief systems on the perception of data rigor. Not only do different stakeholders support different data sets or interpretations, but their support is often based on subjective grounds rather than objective ones. It is not easy to separate our gut feelings or personal preferences from what good data suggest. Everyone has an agenda regardless of whether he works in the private sector, government, or academia. But position or job title is not automatic qualification of one's conclusions or statement of belief as the best available science. Asking hard and pointed questions about the relevance of data or conclusions to the decision to be made is necessary to identify the best science to be had. Probing for support of the relevance will frequently yield more insight into the objectivity or subjectivity of the presenter.

Another relevance factor not always noticed are those data *not* presented as part of the best available science. If it is known that industrial operations have reporting requirements to regulatory or resource agencies that oversee their activities, then it is responsible to look for those data in the studies and reports that are presented as the best available science.

Even with careful attention to all of the above, data relevance is a difficult concept because natural ecosystems are highly dynamic and highly variable. A period of drought or high precipitation will yield data and results that may not be relevant under more normal weather patterns. Rare events really stir things up. Consider the effects of the Mt. St. Helens eruption in 1980 on the structure and function of terrestrial and aquatic ecosystems across a wide geographic area. More recently, the summer of wild fires in Yellowstone National Park in 1988, and heavy flooding in the lower Columbia River basin in February, 1996 all caused major changes in river morphology, sediment distributions and hydraulics. In the latter situation. using pre-flood sediment chemistry to explain post-flood conditions will most likely not be relevant. An illustrative situation of data relevance came up during NPDES[2] permitting discussions with the Oregon Department of Environmental Quality (DEQ) for an industrial site along the lower Willamette River. The company operating on that site was being asked to meet water quality standards of unspecified origin. When the DEQ staffer was asked for the reference site upon which the criteria were based he responded that it was at a bridge 2.5 river miles up from the location of the facility. Further questioning revealed that the agency did not know if the data they were using as a standard were collected during the summer, autumn, winter or spring; whether the samples came from surface water, 0.6 percent of the water column depth or an integrated sample from the entire water column; whether the samples were collected from mid-river or adjacent to

[2]National Pollutant Discharge Elimination System regulating point-source discharges to receiving waters.

the bank (and, in the latter case, which river bank). In other words, they had a set of water quality numbers but knew nothing about what those numbers represented. Continued discussions led to a set of realistic and practical criteria appropriate to the operation and location being permitted.

Question 5: Do I understand the relationship of all data presented to the question under consideration?

Question 6: Are there other potential data sources that have not been referenced or used?

Question 7: Am I comfortable that the conclusions are reasonable and based on the data presented, and that extrapolations are cautions, limited in extent and reasonable?

5 Conclusions

While there are no national or uniform standards to determine the best available science, answering a set of seven questions about three data categories (quantity, quality, and relevance) allows policy-makers and regulators at all levels of government to assess data suitability to the decision they need to make. Each decision is unique to a policy, permit, location, and time so the definition of best available science varies; that is, the context relative to the decision determines what available science is the best. The approach presented in this white paper can be consistently applied while producing situation-specific results. The process and its specificity demonstrates technical soundness and increases the legal defensibility of the decision and the process used to reach it.