Maximizing The Return on Your Environmental Data Investment*

Data Management Best Practices Session
NEBC Business & The Environment
December 11, 2018

Richard B. Shepard, PhD
Applied Ecosystem Services, LLC
Troutdale, OR

Introduction

As business operators, regulators, and attorneys you do not need to be environmental scientists to understand the importance of environmental data relationships to economic, natural, and societal ecosystems. You need to be experts in your professions, not in ecology or environmental science. This brief presentation will help you to appreciate that environmental data are an investment whether you are a business owner, executive, or manager; an environmental regulator; or an environmental, natural resources, or water law attorney.

There are three major components of environmental data management best practices:

1. Acquiring, storing, and administering environmental data. This is not only because your permit issuance or compliance requires them but because they are essential to understanding relationships between a regulated activity and the environments in which it operates.

2. Analyzing and interpreting environmental data. This is arguably the most important, and too frequently improperly done, component. Statistical analyses of geochemical and biological data transform them into useful information. This process provides insights into causality, spatial and temporal patterns and trends, and separates inherent natural variability from anthropogenic changes.

3. Applying established geochemical and ecological theory to interpret analytical results produces knowledge. Using this knowledge allows you to make better informed, technically sound, and legally defensible decisions that accommodate economic, natural, and societal interests.

These three components of environmental data management best practices are especially important in urban environments. Addressing, avoiding, and resolving water quality concerns in urban environments is more complex than in suburban or rural environments when concerns involve federal

laws such as the CWA and Superfund. Urban environments have more impermeable land surface, sources of pollutants and contaminants, complex surface flows, and waste water/sewer piping systems which makes identifying cause-and-effect more difficult.

This presentation does not address water quality standards; that is an important issue for another place and time. Because environmental concerns involve geochemical and biological data with only rare exceptions, proper data management and treatment are the foundation of successful business, regulatory, and legal processes.

The need for environmental data management goes beyond "the regulator made me do it!"[1] Federal environmental laws were created to provide pragmatic balance among the competing priorities of economic, natural, and societal ecosystems. When a balance is achieved among these systems everyone benefits and prospers.

Acquiring, storing, administering environmental data

Acquisition

Environmental data differ from business, financial, economic, and social data in many ways. One of the most important is that all environmental data are unique to a specific location and time; their spatio-temporal signature. This information must be recorded with the collected data if it is to be useful and not a waste of effort and money.

A fundamental law of nature is that time is one-directional. If we do not collect all pertinent data while at a sampling location we cannot go back and collect the missing data later. This is particularly true for aquatic ecosystems, especially streams and rivers. The water sample taken back to the analytical chemical laboratory is no longer representative of the collection site; that water is far downstream by then. This means that the data collection program must be designed to answer questions posed by statute and regulation, such as,

- What is the status of the ecosystem now?
- How might a proposed or existing operation or project adversely affect the ecosystem?
- What do we need to understand about the ecosystem's dynamics to answer the above two questions?

Data collection locations and sampling frequency need to be identified based on how they contribute to answers. A critical issue almost never considered when designing a baseline or permit compliance monitoring program is being able to separate inherent natural variability from permitted discharge from the operation. Requiring permit holders or project proponents to collect extensive baseline and frequent monitoring data is a time and financial burden on everyone and is not needed when the monitoring program is based on sound ecological and statistical theory and understanding. And there is almost always sufficient existing data that, collectively, provide insights on ecosystem dynamics including variability.

Storage

It is an unfortunate reality that too few environmental permit holder and regulatory agency properly stores their environmental data. Spreadsheets are almost universally used and they are the wrong

[1] With apology to Flip Wilson's Geraldine.

tool. Computer spreadsheets are the digital version of paper bookkeeping journals and ledgers. They are great for analyzing financial data but not for storing (or analyzing) environmental data.

Environmental data analysts commonly spend weeks or months wrangling data (as much as 70% of a project's time and expense). This wrangling involves cleaning spreadsheet data, exporting it, translating it to a usable format, and loading it into a database. There are many reasons why spreadsheets are wrong for environmental data. A few of these are:

- They allow text in columns that are supposed to contain numbers. It is not possible to sum a column of numbers in which some rows contain a non-digit such as the less-than symbol, "<."
- They do not prevent entry of the same string in different formats; for example an address in one cell is "909 N. Hayden Island Dr." while another cell has "909 North Hayden Island Drive." A search using one string will not find the cell containing the other string.
- They do not prevent duplicate or invalid data.

Databases prevent the above errors and allow data to be separated by category and stored in only one place. For a water discharge permit compliance monitoring database the database should contain separate tables for data collection sites, surface water chemistry, ground water chemistry, biota (when collected or counted), and bacteria. These tables can be joined in multiple ways to extract data that address questions asked of them.

Setting up and administering an environmental database is as important as setting up and administering a business' bookkeeping and accounting system. The data are an investment in your company or agency and need to be properly cared for in order to gain the benefits they offer.

Administration

Environmental data need attention. This is not complicated or time-consuming but necessary. For example, when an analytical chemical laboratory report is returned it should be carefully read for typos, apparent anomalies (such as a very high concentration), or indicators that the operation might have a water quality issue developing. Anything that looks unusual or questionable should be brought to the lab's attention so they can explain or correct it.

When preparing compliance monitoring reports for regulators describe the operation's water discharge in terms of the designated beneficial uses. This means providing insights into whether the discharge has had an effect on water quality, fish, or wildlife. A spike in concentration of a water constituent might be a periodic response to season, current velocity, or aquatic animal life cycle stage. An example is warm summer temperatures in the mid- to lower-reaches of a river system which occur when anadromous fish are not present and the only resident fish are those that tolerate summer water temperatures. Raw data and statistical time series and trend analysis results should be included in the report while the focus is on answering the question, so what?

Analyzing environmental data

The focus of environmental data analysis should be determining causality and trends; that is, why we observe the values we do and defining how our observations vary spatially and temporally. This is why data collection programs should be designed and used with these goals in mind. To make

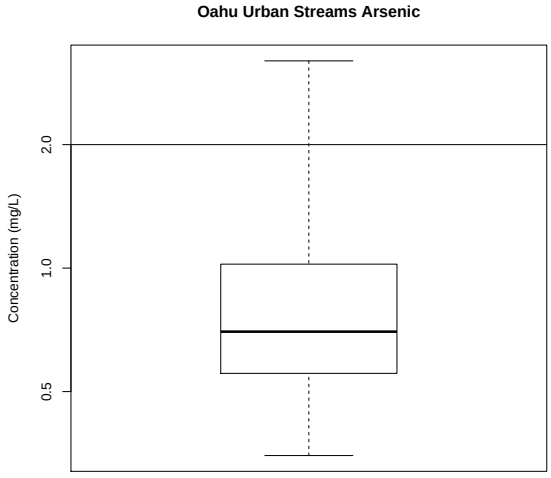

Figure 1: Arsenic concentrations in urban streams on Oahu Island, Hawaii.

informed business, regulatory, and legal decisions we need to collect data on potential explanatory variables as well as the response variable that usually is represented by a regulatory MCL.

Environmental data, both geochemical and biological, are different from other types of data and are particularly sensitive to the assumptions of the models used to analyze them. It is necessary to fit analytical models to the existing environmental data rather than trying to fit the data to a fixed model. Spreadsheets are not appropriate for environmental data analyses, even for descriptive summaries such as average and variance. Environmental data, particularly geochemical data are not normally distributed but the models used in spreadsheets assume the data are.

Geochemical data cannot have negative values, unlike financial or physical data such as temperatures. The geochemical data most commonly of concern are toxic metals and organic compounds. These chemicals tend to be more frequently measured at low concentrations, if they can be detected at all, and may have a few very high values. The better descriptors of geochemical data are the median (the value that has the same number of greater and lesser values in the sample) and the range between the 25^{th} and 75^{th} percentiles (the interquartile range).

Figure 1 illustrates the above points. It shows the distribution of arsenic concentrations in urban streams on Oahu Island, Hawaii. The horizontal line is the analytical chemical laboratory's method detection limit of 2 mg/L. More than 90% of the collected samples were below that value (could not be detected), yet the shape of the distribution of values can be determined with the proper statistical model. The box represents the interquartile range, the dark horizontal line in the box is the median value, and the short horizontal lines at the ends of the dotted "whiskers" are the total range of values. This figure presents a lot of information useful for businesses, regulators, and the public.

There are many other characteristics of geochemical data that should be incorporated into analyses that result in technically sound and legally defensible conclusions.

Using environmental data to support business, regulatory, and legal decisions

Properly acquiring, storing, administering, analyzing, and interpreting environmental data costs time and money so you want to use it most effectively to maximize the return on your investment. Businesses and regulators have the same need approached from opposite sides; attorneys have slightly different needs from environmental data applicable to their cases. However, all three groups benefit by making more informed decisions based on understanding specific operations, locations, and cases.

Business want to survive, grow, and minimize non-revenue producing costs. Baseline and compliance monitoring data collected in support of operating and discharge permits are as valuable as are market, production, and financial data. When a business plans a new operation site on federal public lands that is considered a major action by the land management agency an environmental impact assessment or statement must be prepared to comply with the National Environmental Policy Act (NEPA). The baseline data should both fully describe the dynamics of the ecosystem and propose how the project will be designed and operated to avoid, minimize, or mitigate any adverse impacts. Too often the EA/EIS document is not written with this need in mind and the result is administrative delays and potential legal challenges. When properly done, interpreted baseline data results provide the standard applied to further data that allows adjustment when needed to avoid environmental issues. Technically sound and legally defensible environmental data collection, analysis, and presentation also helps avoid and effectively respond to lawsuits.

Regulators have two goals for the environmental data they collect: demonstrate to the legislature that they are protecting natural environments in accordance with statutes and avoid lawsuits by project objectors such as NGOs. This requires the agency to look beyond single constituent concentrations compared with MCLs to an understanding of the ecosystems for which they are responsible. Regulators also should support environmentally-responsible operations by businesses because it is tax revenue generated by business and their employees that funds agency staff and programs. Same vested interests as business but from the other side of the table.

Attorneys assisting clients or litigating on their behalf need to understand the science involved in the issue being addressed or that underlies the data being litigated. The science can be complex but explained in nontechnical terms so they can better serve their clients and so that finders of fact can make decisions based on understanding the technical facts rather than on emotions.

—

Applied Ecosystem Services, LLC / 2404 SW 22 St, Troutdale, OR 97060-1247 / 503-667-4517 / rshepard@appl-ecosys.com (rev. December 11, 2018)