2004

The Triad Approach: A Catalyst for Maturing Remediation Practice

Deana M. Crumbling
U.S. Environmental Protection Agency

Joel S. Hayworth
Hayworth Engineering Science, Inc.

Robert L. Johnson
Argonne National Laboratory

Marlene Moore

The Triad Approach: A Catalyst for Maturing Remediation Practice

Deana M. Crumbling

Joel S. Hayworth

Robert L. Johnson

Marlene Moore

Many individual scientific and technical disciplines contribute to the multidisciplinary field of remediation science and practice. Because of the relative youth of this enterprise, disciplinary interests sometimes compete and conflict with the primary goal of achieving protective, cost-effective, efficient projects. Convergence of viewpoints toward a more mature, common vision is needed. In addition, cleanup programs are changing under the influence of Brownfields initiatives and the needs of environmental insurance underwriters. Investigations and cleanups increasingly need to be affordable, yet transparent and defensible. Disciplinary goals and terminology need to better reflect real-world site conditions while being more supportive of project needs. Yet, technical considerations alone will not ensure project success; better integration of human factors into project management is also required. The Triad approach is well placed to catalyze maturation of the remediation field because it emphasizes (1) a central theme of managing decision uncertainty; (2) unambiguous technical communications; (3) shortened project life-cycles and multidisciplinary interactions that rapidly build professional expertise and provide feedback to test and perfect programmatic and field practices; and (4) concepts from "softer" sciences (such as economics, cognitive psychology, and decision theory) to capture important human factors. Triad pushes the cleanup industry toward an integrated, practical, second-generation paradigm that can successfully manage the complexities of today's cleanup projects. © 2004 Wiley Periodicals, Inc.

INTRODUCTION

This article will explore how the Triad approach can spur evolution of remediation science and practice from a patchwork of weakly associated disciplines toward a fully integrated multidisciplinary field of scientific inquiry and engineering practice grounded in negotiated, cooperative decision making. The Triad approach was envisioned as a self-correcting paradigm that can keep pace with advancing science and technology as they interact with evolving societal demands. To meet this ambitious goal, Triad is built on a central, powerful, organizing principle: the explicit management of decision uncertainty. Triad was articulated by a team of effective, experienced practitioners with a proven track record of successfully managing complex sites. Their lessons learned and expertise are captured in the workings of the Triad approach. Triad crystallizes this expertise into a second-generation paradigm of integrated practices that can benefit the consumers of site remediation services. Triad also promotes deeper collaboration among field practitioners, regulatory policy makers, technology developers, and academia as it encourages open communication. It feeds on process and technology improvements flowing from both the public and private sectors.

Published online in Wiley Interscience (www.interscience.wiley.com). DOI: 10.1002/rem.20029

RAMIFICATIONS OF UNCERTAINTY MANAGEMENT

Building Triad on the unifying concept of uncertainty management has many ramifications at project and program levels. One ramification is reliance on a flexible, graded approach that can be tailored to meet each project's specific technical needs. A second implication is Triad's openness to advancing science and technology. Triad welcomes any and all tools that offer more accurate and complete understanding of contaminant distribution and behavior, since these are key factors determining the risk of exposure and cost-effective mechanisms for risk management. Another ramification is that Triad's structure allows it to automatically adapt even though the nature of environmentally related decisions may change in response to evolving social, policy, or regulatory initiatives. For example, Brownfields initiatives shift focus from legal liability to economic revitalization, yet the basic need remains for all decisions (whether engineering, legal, or economic) to be protective, transparent, and technically defensible.

Explicit management of decision uncertainty serves the goal of defensible, transparent science in site cleanups.

Explicit management of decision uncertainty serves the goal of defensible, transparent science in site cleanups. The hallmark of scientific inquiry is identifying and controlling (to the extent possible) the variables that could confound the interpretation of data and lead to faulty conclusions. Like all science-based strategies, Triad continually tests assumptions, even long-standing ones, for validity in the face of new information. This is what makes science a self-correcting paradigm. Uncertainty management provides a conduit for feedback between field practice and program policy. Programmatic policies about data quality and statistics developed back in the 1980s may be inadequate now that there is better understanding of the physical mechanisms governing heterogeneous contaminant concentrations in environmental media (Crumbling, 2002; Interstate Technology and Regulatory Council [ITRC], 2003). Program-level policies and procedures that ignore major sources of scientific uncertainty at the project level abet project inefficiency. Sweeping significant uncertainties under the rug does not cause them to go away. They lurk below awareness to trip up the unwary and induce faulty conclusions about risk or remedial design. Better communication between practitioners and policy makers about what really works and what does not in technical practice is needed.

Project success also depends on managing non-science uncertainties. Triad's broad interpretation of "decision uncertainty" creates ample space within Triad's first element (systematic planning) to accommodate the many non-science issues that impact site cleanup, such as uncertainty about budgets and contracts, stakeholder interests and fears, legal concerns, and regulatory interpretation. All these uncertainties affect how a project's end goals are framed, shaping the decisions that must be made to bring the site to closure and reuse. Seasoned Triad practitioners recognize that the people-oriented aspects of a project are as important to success as the scientific and technical. People issues can make or break a project.

Starting fieldwork before there is consensus about the desired project outcome is one reason why repeated field mobilizations are standard fare for conventional projects. Making room in up-front project planning to confront regulatory and community concerns head-on is a major reason why Triad projects move rapidly and efficiently during field implementation. How long it takes to get to the field, however, depends on the willingness of those involved to work toward consensus during planning. For some projects, planning is relatively quick and simple because the issues are clear-cut and participants are few or are motivated to reach consensus. For other projects, just articulating what the project goals are (much less a workable strategy to reach them) can be a difficult and lengthy process, with divergent stakeholder interests pulling in opposite directions. Triad cannot miraculously change participants' basic motivations. It can, however, create a forum where all concerns and suggestions are openly vetted, valued, and factored into the decision process. Honest communications, transparency, and accountability can sometimes work wonders for contentious projects. As difficult as Triad planning can sometimes be, it is still much more cost-effective to resolve conflict and uncertainty about project goals through face-to-face meetings, rather than hoping to satisfy stakeholders through a succession of trial-and-error work plans and repeat mobilizations (Crumbling et al., 2004).

Triad's core principle catalyzes better projects because Triad practitioners use decision uncertainty management to bring order to an otherwise chaotic list of seemingly unrelated or contradictory activities and issues. Topics ranging from data quality, remedial optimization, and long-term monitoring to stakeholder involvement, insurance underwriting, legal defensibility, and procurement can be aligned and harmonized when they are subordinated to serving the same "master"—i.e., achieving confidence in the key decisions that determine the project outcome. The same harmonization can be applied at the program level to ask whether established procedures are serving the overarching goal of decision confidence.

The Triad approach can benefit the cleanup community in yet another way. Triad and its core organizing principle can help catalyze the emergence of a fully integrated site management discipline from the loose confederation of disciplines contributing to environmental investigation and cleanup. The multidisciplinary heritage of remediation science is rich and diverse. Each discipline makes a vital, unique contribution. But a unified scientific discipline needs a common vision and a common language in order to progress and thrive. Right now, remediation science has neither. Each distinct discipline has retained its own narrow view of what constitutes good practice in its particular realm. Each discipline retains unique jargon despite the miscommunications it causes. Independent pursuit of parochial interests has worked against the overarching goal of achieving quality at the project level, where science-based decisions about exposure and remediation must be made at the intersection of social concerns and economic interests. In hope of sparking a dialogue leading to greater harmonization of multidisciplinary efforts, two disciplines—analytical chemistry and statistics—will be used to illustrate how narrow disciplinary frameworks work against project efficiency and defensibility. Then the article will explore how decision theory aspects of the Triad approach deepen productive collaboration between disciplines to promote continual improvement and professional development.

Honest communications, transparency, and accountability can sometimes work wonders for contentious projects.

ANALYTICAL CHEMISTRY

From the standpoint of analytical chemistry, data quality for chemical pollutant analyses is a function of analytical rigor and instrumentation, analytical quality control (QC), reporting, and review. This thinking permeated the remediation arena in the earliest days of site investigation and has persisted to this time. Throughout the environmental field, the phrase *data quality* is used universally when only analytical quality is being evaluated. The assumption is that the better the quality of the analysis (i.e., precise, unbiased, fully documented laboratory procedures), the better the data quality. Yet the term *data quality* is

Many bench chemists view detecting and controlling the impact of heterogeneity on data quality as someone else's problem.

also identified intuitively and explicitly with the usability of data to make decisions (US EPA, 2000b). Because language has equated analytical quality with data quality, the working model is "good quality analysis = good quality data = good decisions."

Although seductive in its simplicity, the problem is that this model doesn't work very well for contaminated sites. Good quality analysis routinely provides data that leads to erroneous conclusions about the nature and extent of contamination, to the detriment of project defensibility, efficiency, and cost. Why? Although there are several mechanisms by which this happens, the shortest explanation is that contaminated media are heterogeneous at both larger (macro, *between* samples across the site, at the scale of project decisions) and smaller (micro, *within* a single potential sample, at the scale of sample analysis) spatial scales. It is not unusual for both types of heterogeneity to be severe (ITRC, 2003; US EPA, 2003). Since *data* are generated from heterogeneous *samples*, the effect of sampling variability must be taken into account when assessing data quality (i.e., the ability to trust that the data lead to correct decisions). Sampling variability can cause soil or groundwater samples from the "same" location or well to provide radically different results as a function solely of how sample collection and processing were performed. This effect is distinct from analytical problems caused by matrix interferences, which can also occur. Although data variability due to sampling varies widely with matrix, analyte, and estimation method, it usually overwhelms analytical variability in site data sets. Rigorous, quantitative analyses are important and valuable; however, accurate laboratory analysis by itself is not sufficient to cultivate efficient projects. Accurate analyses on the tiny samples actually extracted for laboratory analysis do not guarantee good data quality if there is no confidence that the results can be reliably extrapolated back to the volume of site matrix being targeted by project decisions.

Many bench chemists view detecting and controlling the impact of heterogeneity on data quality as someone else's problem. But the pervasiveness of sampling variability makes it everyone's problem. Everyone, from policy makers to project managers to field technicians, needs to be educated about how heterogeneity sabotages *data* quality (and project success) despite regulatory efforts to ensure *analytical* quality. By colluding with the "analytical quality = data quality" fallacy, chemists have allowed data users to remain ignorant of, and evade responsibility for, other components of data quality. While significant resources are channeled into extensive measures to oversee select portions of analytical quality, much larger sources of data variability are neglected. Ironically, the very tools that can cost-effectively detect and manage the impacts of sampling uncertainty (such as inexpensive, high-density, rapid turnaround field and lab techniques) have been discouraged by regulators and many bench chemists. Entrenched terminology disparages *any* analysis done in the field as "field screening" (Crumbling et al., 2003). For example, field-portable gas chromatography-mass spectrometry (GC-MS) results with a full quality control package demonstrating performance equivalent to or better than fixed laboratory data have been rejected by regulators who identify the technique as "field screening." The term *field screening* implies something less than adequate for making decisions. The reality is since the method performance is equivalent and more samples are measured in less time with the additional benefit of sample contaminant stability, the information better represents the site condition. This attitude toward field measurements has retarded practitioners' and service providers' opportunities to learn how to use and deploy these techniques effectively.

Paradoxically, bench chemists bemoan data users' unfair criticism of the lab when data "doesn't make sense" because of sampling variability that was out of their control.

Chemists can correct this by changing their terminology so it reflects the project perspective and educates data users about data uncertainty. Chemists should stop using the broader term *data quality* and replace it with the more precise term *analytical quality*, since that is almost always what is actually being considered. *Data quality* should be restricted to usage where both analytical and sampling uncertainties have been controlled. Control over sampling uncertainty is as much or more of a contributor to data quality as analytical control. Even perfect *analytical* quality (at the scale of analysis) should be labeled as "screening quality *data*" if uncontrolled sampling uncertainty compromises confident extrapolation of results to larger volumes of matrix (at the scale of the decision). Chemists should shun the custom of analyzing a single soil subsample from a jar and reporting a result to one or two decimal places, when it is well known that other subsamples from the same specimen jar can produce very different results because of micro heterogeneity (ITRC, 2003). Terminology usage and reporting conventions allow data users to remain unaware that "data" validation/verification procedures do little to assess data uncertainty. Data users and project planning teams cannot learn to avoid data errors until the community escapes from the "analytical quality = data quality" language trap.

STATISTICS

The first-generation model the US EPA chose to use for assessing decision confidence for contaminated sites was classical statistics. Since classical statistics are still being promoted as a quantifiable, objective means to demonstrate decision confidence, it is important to consider:

1) Are classical statistics the proper statistical model for routine application for contaminated site projects?
2) Will promoting statistical software packages to inadequately trained staff for planning sampling designs and evaluating data help or hurt the cleanup community?

Are Classical Statistics the Appropriate Model?

The Achilles heel of any statistics application is failure to define the population of interest and demonstrate that data were actually drawn from that population. Good and Hardin (2003) warn that "extrapolation from a sample or samples to a larger incompletely examined population must entail a leap of faith." (p. 3). For lack of a better model, the US EPA has long made "a leap of faith" in recommending classical statistical algorithms for planning sampling and for evaluating contaminant concentration results for hazardous waste sites (US EPA, 2000a, b). These recommendations encourage practitioners to think there is only a single population enclosed within the site boundaries. Yet the physical mechanisms of contaminant release and migration guarantee that site boundaries enclose two or more different contaminant populations whose distribution patterns have strong spatial correlations. Physical reality at most sites violates the model assumptions of classical statistics, and mismatches between the scale of decision making and the scale of analysis are not addressed.

The Achilles heel of any statistics application is failure to define the population of interest and demonstrate that data were actually drawn from that population.

Disturbingly, our experience with the cleanup community has provided ample evidence that project staff routinely use statistical tools without awareness of model assumptions. It is rare to find that staff entrusted with statistical analysis actually under-

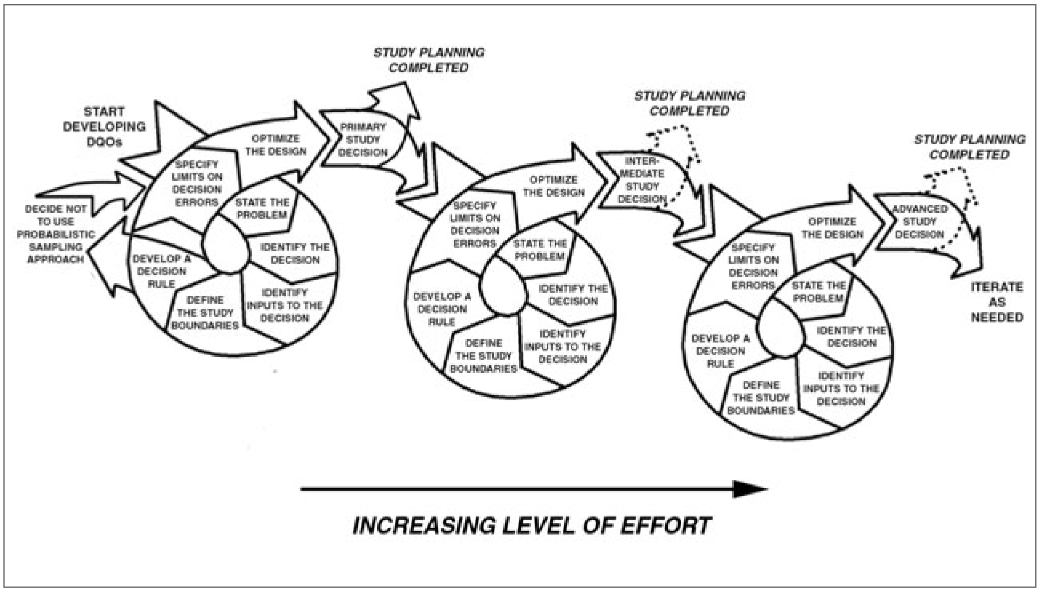

Exhibit 1. Figure 0-4 reproduced from page 0-8 of US EPA QA/G-4 guidance, entitled "Repeated Application of the DQO Process throughout the Life Cycle of a Project" (US EPA, 2000b).

stand the concept of "population" and how to use it to improve project efficiency. The repercussions of using unrealistic values as inputs to classical statistics algorithms to prepare sampling plans (e.g., the population standard deviation and setting the "gray region" as defined by US EPA guidance) are seldom appreciated. Although recognition is growing that geostatistics is a more appropriate statistical model for spatially patterned contamination, classical statistical equations are tightly embedded in practice and procedure. There are good reasons to ask whether classical statistics should continue to be encouraged as the primary statistical paradigm for dealing with contaminated sites:

1) The history of this paradigm is that several site characterization mobilization cycles are required to answer project questions, and even then, the characterization is often discovered later in the project to be flawed, requiring it to be done yet again. Could it be that classical statistics is simply not the right tool if we expect projects to get the "right answer" the first time?

Working within the framework of the time, the US EPA's Data Quality Objectives (DQO) process was designed to reflect the conventional phased approach of fixed laboratory analysis with multiple site mobilization phases. The DQO process was envisioned as a looping of multiple study events, as illustrated in Exhibit 1 (US EPA, 2000b). Although there is no absolute requirement that these study events be separated in time, the expectation was that multiple mobilizations would iteratively hone in on the right answer. This was a reasonable approach when tools, experience, and knowledge were limited, yet social priorities devoted large budgets to site management. The statistical DQO model was not designed to support site characterization that gets the "right answer" in only one or two mobilizations because cleanup programs did not have that expectation—the theory and technology to do so did not exist at that time.

Times and circumstances have changed. The phased mobilization and decision-making model is no longer viable in today's budget climate. This is especially true for Brownfields projects, which have neither the funding nor the time to support multiple attempts at contaminant characterization. Once the budget allotted for characterization is expended, project decisions will often be made whether data support confident decisions or not. Unfortunately, a "best guess" based on incomplete or misleading data sets risks project overruns or outright failure during remediation or redevelopment. Today's leaner budgets and time-critical reuse scenarios require that iterations to find and fill data gaps be performed through real-time in-field decision making if the "right answer" is to be obtained cost-effectively. Fortunately, technical advances make that option possible. However, regulators and practitioners must be willing to move beyond first-generation models and invest in the changes needed to adopt another model. Although nothing in DQO guidance inherently prohibits it, the environmental community has shown no inclination to use the DQO process as a springboard to a second-generation paradigm. To the contrary, many claiming to use the traditional DQO process seem reluctant to break away from their comfort zone, even while they complain of the current framework's inefficiencies.

2) We should not be surprised that classical statistical tools require multiple iterations to get the right answer since fundamental assumptions of the model are incompatible with real conditions at contaminated sites, as described below.

- Statistical models assume that the user has established the validity of the inputs before using them in the model. As one observer noted, "Fancy statistical methods will not rescue garbage data" (R. J. Carroll, 2001, in Good & Hardin, 2003, p. 25). The model assumes that data inputs are fully representative of the population of interest—i.e., that analytical or sampling variability has been controlled. But, as discussed earlier in this article, that is not a safe assumption for environmental data. The concept of population for soil or groundwater is somewhat different from the traditional statistical concept of population, which entails grouping *individuals* with similar characteristics (like a population of maple trees). Extending classical statistics to populations without discrete, clearly delimited individuals or members introduces important theoretical and practical issues. Since soil and water do not exist as obvious collections of discrete individuals, the data user is required to define the population of interest before sampling occurs. Unless sample collection and analysis have been carefully planned to target the population of interest to the decision, traditional sampling and analysis procedures can unknowingly mix different populations together to create misleading intermediate results (an important problem in groundwater monitoring) or can unwittingly target nonrepresentative populations (such as an inappropriate particle size during laboratory subsampling of soil). Undue trust in statistics without a healthy skepticism about the validity of the inputs risks faulty conclusions.
- Classical statistical equations are based on the assumption that each data point is independent. In other words, a fundamental assumption is that there is *no spatial relationship* between the data points. Another way to phrase this is that the data are expected to come from a single, identically distributed population. That is why classical statistic equations do not consider the actual area of the site or the volume of matrix when predicting the number of samples to be collected. To classical statistics, the area or volume being sampled is irrelevant; therefore, it will tell you to take the same number of samples whether you are sampling 1 acre, 100 acres, or 1,000 acres. Obviously, no one believes this. The reason for disbelief is that disposal practices and contaminant migration are known to create

Undue trust in statistics without a healthy skepticism about the validity of the inputs risks faulty conclusions.

spatial patterns that limit our ability to extrapolate results across large areas. The degree of limitation is related to the representativeness of sample results and is determined in Triad practice through the process of refining the conceptual site model (CSM) and managing the relationship between decision uncertainty and data uncertainty (Crumbling, 2004; ITRC, 2003). Most project decisions related to assessing exposure pathways and designing cost-effective remedial systems are dependent on knowing the spatial distribution of contaminants. Those decisions cannot be properly made if supporting data were produced through a sampling strategy that assumed no spatial relationship existed.

  - Note that the assumption of a single population may hold true for select scenarios: (1) sites that were never contaminated; (2) sites that have been successfully cleaned up; (3) sites that were contaminated through some unusual mechanism that uniformly covered the entire site; and (4) stratification or blocking strategies that allow sections of a site or specified volumes of matrix to be defined as a single population because there is reason to believe contamination is reasonably homogeneous within the specified boundaries of that population.

Our experience with users of statistical sampling design tools is that they consistently underestimate the degree of real-world variability.

- Algorithms to support sampling design development presume a level of site understanding (e.g., predicting the variability in the future data set) that usually does not exist. This requires that guesses or professional judgment be used as inputs to the equations. Our experience with users of statistical sampling design tools is that they consistently underestimate the degree of real-world variability. The concept of the gray region is not understood at all. Sampling programs that attempt to treat a large site as a statistical whole invariably oversample in some areas but undersample in others, creating an inefficient data set with gaps that trigger additional mobilizations.
- Many applications of classical statistics are structured to assume that the project decision rests on a comparison between the population mean and a regulatory threshold. For several reasons, this assumption is frequently violated for contaminated site projects:
  - Even for surface soil contamination (the scenario most often depicted in guidance and training examples), current regulatory compliance decisions are seldom based on areawide averages. Regulatory thresholds are almost always treated by regulators as "never-to-exceed" levels. Even if the areawide estimate of the average (the upper confidence limit [UCL] on the mean is most commonly used) is below the regulatory limit, regulators will commonly insist that data points exceeding the threshold be investigated by another round of sampling. Regulators intuitively recognize that significant masses of contamination or migration pathways can be missed by widely spaced sampling designs. Even one "hit" above the limit could represent the "tip of the iceberg" if the original sampling plan was not dense enough to control for spatial heterogeneity or did not delineate high hits to determine extent. However, treating regulatory thresholds as "never-to-exceed" levels creates its own host of problems from the standpoint of scientific and regulatory defensibility.
  - As noted before, a data set representative of a sitewide (or large areawide) average is not useful for making the many project decisions that

depend on detecting concentration gradients or spatial patterns. Examples of these decisions include detecting small cross-section/high flux migration/exposure pathways, finding sources, selecting and designing an efficient remedy, projecting redevelopment costs and schedules, and pricing environmental insurance policies.
    - Traditional project planning seldom describes the relationship between contaminant populations and the intended project decisions; therefore, sample collection and processing is not planned to be representative of the population targeted by the intended decision. The result is haphazard sampling and sample processing that easily produces a data set unknowingly drawn from different populations. The mean of a data set that unintentionally and unknowingly mixes different populations in the data set is unreliable as a basis for decisions using the mean.
- Classical statistics themselves provide no mechanism for considering information sources other than chemistry results for pollutant concentrations. Users typically employ statistics as if the chemical data set is expected to stand alone to support decisions. As already discussed, that is a risky proposition when target populations are poorly defined and sampling variability is uncontrolled, making sample representativeness highly uncertain. Other data sources (e.g., geophysical survey results, stratigraphic data) and the site history are crucial to establish the physical context in which chemical data should be interpreted. Innovative algorithms based on Bayesian statistics have been used to allow the site's physical context and other knowledge to be included quantitatively when estimating decision confidence (US DOE, 2001).
- Chemists and statisticians sometimes recommend that data users employ statistical outlier tests to tidy up data sets and justify discarding inconvenient data points. This practice can actually work against project success. Chemical data sets should not be treated as stand-alone systems. Chemical data reflect the physical nature of a site, a nature that is governed by interactions that are best understood within the context of other technical disciplines such as geology or soil science. Discarding data based on a statistical model may be throwing away valuable clues to understanding the site. Results that appear to be outliers may be important clues that the site does not conform to the statistical assumption of a single population. Outliers can also be telltale signs that the working CSM is not correct, that target populations have not been correctly identified, or that sampling uncertainties are inadequately controlled. Rogue results are warning signals that, if heeded, allow problems to be detected early so corrective actions can be taken, avoiding more serious and costly difficulties later in the project. Outliers should be discarded only when they can be traced to a blunder or have been replaced with more solid information.

Discarding data based on a statistical model may be throwing away valuable clues to understanding the site.

The disconnect between the assumptions inherent to classical statistics and the realities of contaminated site projects should be a cause for concern among statisticians and project managers. Although it is possible to apply classical statistics in ways that preserve the integrity of the assumptions and allow the tool to be used defensibly, our observations are that most environmental users are unaware that statistical pitfalls even exist, much less how to avoid them.

The Perils of Using Statistical Software as a Black-Box

It is vital that the environmental community heed the warnings of statistical experts: "[S]tatistical software will no more make one a statistician than would a scalpel turn one into a neurosurgeon. Allowing these tools to do our thinking for us is a sure recipe for disaster." "Statistical procedures for hypothesis testing, estimation, and model building . . . should never be quoted as the sole basis for making a decision . . . the most serious source of error lies in letting statistics make decisions for you" (Good & Hardin, 2003, pp. ix, 3). Within the environmental community, statistically based programs are frequently used as "black-boxes"; that is, inadequately trained staff run software and accept its output without (1) understanding what the model assumptions are, (2) establishing whether model assumptions are valid for the specific application, and/or (3) determining whether project-specific inputs to the program are justified. For example, inputs to statistical software programs to calculate sample numbers (such as values used for the gray region and the population standard deviation) are often selected without regard for actual site conditions. Since these inputs determine the model's output (how many samples are needed to support decision making), using inputs that are qualitative estimates or guesses at best, or factitious values selected to achieve a pre-determined outcome at worst, produces equally uncertain or factitious outputs. But the sensitivity of statistical models to the validity of inputs is too often downplayed in the policy arena, luring managers and staff into a false sense of security. Passing guesses through a mathematical algorithm does not make the output more "scientific" and "quantitative," despite sweeping claims that using these statistical methods will "design the data collection plan that will most efficiently control the probability of making an incorrect decision" (US EPA, 2000a, p. 5).

Unrealistic proposed sampling designs have been defended simply because a software package was used; the rationale for choice of inputs is not discussed.

This warning applies equally to classical, geostatistical, and geo-Bayesian algorithms. When human statisticians were used to design sampling plans that considered the highly variable nature of site contamination, the number of samples they recommended routinely exceeded the characterization budget. This was a clue that a purely statistical approach to sampling design was inadequate as a design model, but viable alternatives were not available. So project managers did the best they could within budget constraints. They ignored the statistical calculations in favor of calculating how many samples the budget could support and that became the basis for sampling design. If statistical software programs are used, the same problem (conventional statistically based sampling designs are too expensive to implement) should arise, since the software runs the same equations that the statistician did. What black-boxes offer, however, is the ability for a nonstatistician to interactively run the algorithm until the set of input values is found that predicts the number of samples that matches the budget. Although testing a proposed plan against the budget is a legitimate activity, selecting unrealistic inputs defeats the purpose of using statistics, which is to increase the objectivity and defensibility of the plan. Unrealistic proposed sampling designs have been defended simply because a software package was used; the rationale for choice of inputs is not discussed. Case managers providing oversight seldom have the training to detect invalid statistical inputs. Reliance on black-box statistics will keep the environmental community trapped in the paradigm of multiple mobilizations that resample until the budget runs out, settling for an incomplete CSM that sabotages any chance for efficient remedial and monitoring designs.

TRIAD AS A CATALYST FOR CHANGE

The environmental community needs to refine its thinking about data quality and statistics to match our current understanding of contaminant heterogeneity and its impacts on decision confidence. Better site characterization would result, producing a much more accurate picture of site contamination. Projects would be more successful, and remediation science would be put on a more rigorous technical foundation. Triad provides a framework to break away from first-generation practices toward a new paradigm that embraces the tools and strategies proven to give much better project outcomes because they manage heterogeneity. The Triad approach explicitly recognizes that physical mechanisms of contaminant release and migration create patterns of contaminant distributions, incorporating them into a CSM that segregates the site into populations supportive of protective, yet cost-effective, project decisions. Populations can be defined at scales matched to support exposure decisions, risk management strategies, and efficient treatment design. Sampling designs are scaled to match the population targeted by the intended decisions. The availability of affordable, rapid, high-density data collection options often allows Triad projects to generate sampling densities that characterize and delineate populations directly, with less need for uncertain statistical extrapolation. Triad is organized around a form of "hypothesis testing" based in physical reality, rather than in statistical models. Triad uses a preliminary CSM as the initial hypothesis of site contamination and its relationship with its physical surroundings. The accuracy of that physical model is progressively tested and refined by real-time iterations of data collection and CSM updates until there is confidence (which may, as the situation warrants, be expressed statistically or through weight of evidence) that the CSM is accurate enough to support correct decisions. To do this, Triad exploits all available tools, including statistics, in a manner consistent with the need and the tool. There are enough technology options now that the need can drive tool selection, instead of the project being designed around a few available tools.

The language of site professionals needs to evolve past the assumptions of the first-generation model and support cross-discipline communication about the physical nature of contaminated sites. As the interagency Triad workgroup prepared the Triad Resource Center Web site, we discovered we needed to use language more precisely to communicate concepts about data quality and sources of data and decision uncertainty. Common phrases, such as *confirmation sampling*, *false positive/false negative*, *DQOs*, and *source area* routinely caused confusion because they meant different things to different people in different contexts. We found we needed to integrate theoretical and practical considerations when defining terms unambiguously. These definitions are captured in the Triad Resource Center's glossary (www.triadcentral.org).

. . .Triad exploits all available tools, including statistics, in a manner consistent with the need and the tool.

When site cleanup began emerging as a discipline, project managers, often engineers, were expected to serve as the integration hub for the disparate disciplines that contribute to cleanup projects. But this was unrealistic. Engineers are seldom cross-trained in multiple fields well enough to enable them to mold disciplinary interests to serve specific project needs. Specialists are better suited for that task, but they need to understand the big picture to do so. It is difficult for specialists to develop that broader perspective. Few analytical chemists or statisticians have had the opportunity to work on a project from start to finish—from sampling design through data generation to CSM refinement and decision making, exposure assessment, and remedial design. They rarely see how their disciplinary choices impact project efficiency and costs. They are seldom exposed to failed dig-and-haul projects or to inefficient pump-and-treat groundwater remedial designs. Without that exposure, they cannot process what goes wrong and be part of a solution. Statisticians don't find out whether actual site conditions bear any resemblance to the model used to predict sampling designs. Laboratory chemists don't experience the consternation of crippled data sets because diluting out interferences raised detection limits. This means they are unlikely to develop ideas for improving their procedures to better serve project needs. This is partly because the disciplines are "stove-piped," and this is due, in part, to the long time frames of projects. Months and years may pass between project planning, execution, and final resolution. People move on to other projects or other jobs, or simply forget what they were thinking when they developed a plan. Feedback about what really contributes to project success is lacking. This makes it difficult for the environmental community to develop general expertise (Klein, 2003).

As part of its emphasis on the management of decision uncertainty, intense and focused planning is a key Triad activity

Until communication channels are opened and feedback loops closed, remediation science will labor as a hodgepodge of competing scientific interests and inconsistent terminology. More opportunity is needed to exchange lessons learned that broaden specialists' perspectives, especially those who are in a position to update the regulatory and institutional frameworks governing investigation and remediation practices. The Triad approach can help. Triad projects explicitly rely on the multidisciplinary collaboration of "allied environmental professionals" (Crumbling et al., 2003). In addition, the short project life-cycles typical of Triad projects encourage feedback about what works and what doesn't (Crumbling et al., 2004). Close collaboration and rapid feedback allows analytical chemists and statisticians (and others in specialty disciplines) working on a Triad project to see it through planning to decision making and its consequences, to learn lessons firsthand, to confront faulty mental models and assumptions, and to devise better procedures. Whereas isolation allows specialists to cling to narrow interests or standard assumptions (e.g., analytical perfection and assumptions of homogeneity), shared responsibility for project success should motivate them to tailor their contributions to serve the overarching goal of successful, cost-effective projects.

STRENGTHENING THE DECISION-MAKING PROCESS

As part of its emphasis on the management of decision uncertainty, intense and focused planning is a key Triad activity. Triad's critical emphasis on detailed planning seeks to avoid the all-too-common pitfall noted by Good and Hardin: "The vast majority of errors in statistics—and, not incidentally, in most human endeavors—arise from a reluctance (or even an inability) to plan" (2003, p. 25). Triad project planning and execution are structured to exploit powerful decision-making strategies. Each element of Triad involves decision-making processes that range from the structured *rational choice strategy* to the more dynamic *recognition-primed decision model* (Klein, 1999). The rational choice strategy (RCS) can be thought of as the classical decision analysis method. When applying an RCS, the decision maker identifies a set of options and ways to evaluate those options. He weights each evaluation criteria, ranks options with the weighted criteria, and picks the option with the highest score. Listing and counting "pros and cons" is an example of a rational choice strategy. In contrast, the recognition-primed decision (RPD) model is a naturalistic method that combines two decision processes: pattern and cue recognition and mental simulation. In the RPD model, the decision makers use prior experience to jump to the right answer without sorting through all available options as done in the RCS.

Both decision strategies have an important place in the Triad approach. Expert Triad practitioners freely move between both strategies as the project progresses. Understanding how Triad can catalyze better projects and maturation of remediation practice requires understanding how project decisions are first articulated and operationalized during planning, then made during project execution. The rational choice strategy is ideally suited as a decision model for the systematic planning element of Triad. RCS provides a deliberate, quantitative, systematic approach to complex problems. It reduces the chances that an important consideration will be overlooked. It also allows for more in-depth analysis of many options. The RCS protects novices from making poor choices, and is useful when working in teams under minimal time pressure. One can think of the systematic planning process as a shared commitment to "get it right" from a planning perspective prior to committing to expensive field deployment. Using an airplane analogy, the "landing" is planned before the project "takes off" for the field. The RCS decision model gives all participants a sense of order, structure, and confidence that all worthwhile strategies have been considered and contingencies addressed. It also allows participants who may not yet be "experts" to observe and learn from experts while being one of the team and contributing in their own right. This is one of the ways Triad catalyzes professional learning and competency development.

In contrast, the RPD model is applicable in decision-making environments where experienced decision makers coordinate a team under time pressure and high stakes, often under rapidly changing conditions or inadequate information. The decision maker must be prepared to adapt. Emergency personnel and soldiers work under these kinds of conditions. The dynamic field implementation component of Triad also exhibits many or all of these characteristics, so the RPD model is most applicable to that phase of a Triad project.

But aspects of the RPD model also show up during Triad systematic planning, where the art of identifying and planning for contingencies requires pattern recognition (e.g., drawing on past experience to see similarities and differences in sites) and mental simulation (e.g., mentally walking through a scenario to construct contingency strategies to deal with potential obstacles). In turn, a dynamic strategy will not be successful unless RCS-structured planning was thorough in finding the best options given project constraints and stakeholder concerns. So it is imperative that the core technical team of disciplinary experts has the requisite specialized experience and skills. The principal components of the dynamic strategy are codified as decision flowcharts or matrices. These "if-then" tools will guide the decision scenarios during implementation so that the wishes of the planners can be carried out. Done correctly, the output of Triad systematic planning will be a well-crafted dynamic strategy that addresses goals and contingencies, with all participants feeling prepared to execute the plan.

Triad projects compress the standard time frames of remedial investigations by providing flexibility to respond to real-time discoveries.

The decision trees lay out the overall decision strategy for approval by regulators. They will guide how field activities adapt to the continuously evolving CSM, which has been designed to bring actual site conditions into sharp focus, no matter how heterogeneous those conditions turn out to be. Triad projects compress the standard time frames of remedial investigations by providing flexibility to respond to real-time discoveries. This is a dynamic decision environment with extremely high stakes: incomplete information with the pressures of time and budget do not allow decisions to be deferred like conventional projects. One can imagine an analogy to military combat operations: once a plan is under way, it is imperative that the on-site forces be able to react quickly and decisively to a changing context. To the extent possible, the changing context is anticipated during the systematic planning phase, but larger or smaller surprises are a fact of life. Anticipating what might go wrong so surprises will be manageable is the reason why the core technical team is a group of allied professionals representing all relevant disciplinary specialists. The dynamic strategy must be sufficient to provide the field decision maker with guidelines to operate within a changing context. Equally important, the decision maker must be capable of recognizing if the planned strategy is so severely violated by actual site conditions that he needs to "retreat."

Being able to manage the increased data load and extract the information most relevant to a decision point in real time is crucial.

For decision making during a Triad remedial investigation, the most critical RPD skills are pattern recognition (or pattern matching) and mental simulation. Pattern recognition refers to the ability of an expert to detect typical patterns and/or detect anomalies that violate an expected pattern. This is the reason why experienced technical staff are crucial to Triad: experts can recognize patterns that novices may miss, and notice when cues warn that an expected pattern is being violated. Triad projects "grow" experts quickly. As the CSM is tested and refined in real time, practitioners quickly learn what field cues are associated with contamination, for example, where chlorinated solvents tend to migrate or "get stuck" in subsurface stratigraphy, or what chemically stressed vegetation looks like. A significant component of pattern matching is situational awareness; that is, the ability to observe the "big picture," and filter out irrelevant noise. In Triad projects, information throughput is very high. Being able to manage the increased data load and extract the information most relevant to a decision point in real time is crucial. In contrast, the traditional approach, which expects multiple return trips to the field, relegates this function to the office after a field mobilization is complete.

Another important RPD skill is mental simulation. Mental simulation can be thought of as the thinking that allows experts to explain how past events caused the present situation (e.g., how an observed contaminant distribution came to be) and how the present will impact the future (e.g., how an observed contaminant distribution will behave under natural or induced conditions) (Klein, 1999). Exhibit 2 sketches the coupling of pattern recognition and mental simulation that are part of the RPD during Triad investigations.

The rapid feedback provided during Triad execution quickly lets the team know whether a mental simulation was correct, whether the plan is working, or whether something important was overlooked. These experiential lessons are strongly imprinted into a practitioner's skill set, providing fodder for mental simulation in future projects. Even if the first Triad project experience didn't go that well, the next project will benefit from those lessons learned. Through case studies of Triad projects (posted through the Triad Resource Center Web site) and published articles, these lessons will be passed on to other practitioners, helping to raise the overall proficiency of the cleanup industry.

BUILT FROM THE BOTTOM-UP FOR TOP-DOWN SUCCESS

The Triad approach is designed from the "bottom-up" in the sense that it is built on strategies that practitioners use at the "ground level" to achieve successful projects. But institutional barriers to Triad can be mitigated only by reformulating high-level adminis-

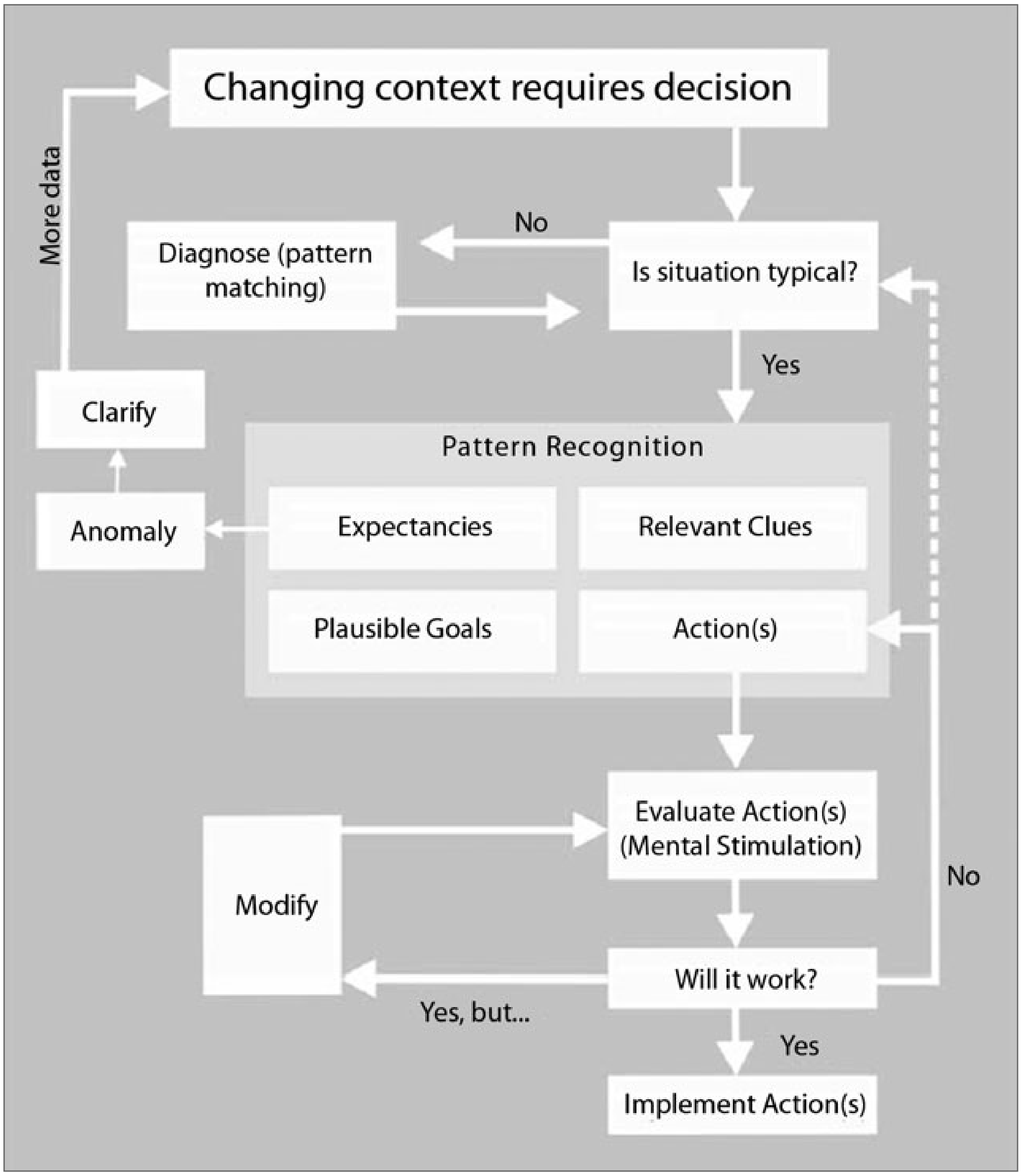

Exhibit 2. Coupled RPD pattern recognition-mental simulation model for Triad projects (based on Klein, 1999)

trative and regulatory strategies of cleanup programs from the "top-down." The success of an administrative program is dependent on individual successes at the project level. Active feedback between the two levels can ensure that programmatic procedures are continually tailored to facilitate project success. Triad concepts and project examples are being used to create a communications bridge between top-down, high-level administrative procedures and the bottom-up technical strategies.

Although Triad is primarily driven by science, it creates ample space for addressing the social, economic, and bureaucratic constraints that strongly influence cleanup projects. Scientific and non-scientific considerations strongly influence each other. Science-based activities cannot be implemented unless higher-level policy and guidance supports the effort. In turn, the economic viability and reputation of cleanup programs and related public policies are dependent on our ability to demonstrate that our technical activities and cleanups actually protect human health and the environment without squandering increasingly scarce resources. Before Triad was called Triad, one of the early proposals for labeling this second-generation paradigm was "the Gestalt Approach." Although the term *Triad* won the vote, the term *gestalt* accurately reflects the holistic, "big picture" framework Triad embraces, where the unified whole is greater than the sum of its parts. Like a fine orchestra, each instrument lends its voice to an interwoven pattern possible only when each player harmonizes with the others while working toward a common goal. The multifaceted, multidisciplinary nature of environmental cleanup requires a paradigm that can catalyze the environmental community to accomplish nothing less.

REFERENCES

Crumbling. D. M. (2002). In search of representativeness: Evolving the environmental data quality model. Quality Assurance, 9(3-4), 179–190. Retrieved October 1, 2004, from http://cluin.org/download/char/dataquality/dcrumbling.pdf

Crumbling, D. M. (2004). The Triad approach to managing uncertainty in environmental data. NEMC 2004 Conference Proceedings. Retrieved October 1, 2004, from http://www.triadcentral.org/ref/room/index.cfm

Crumbling, D. M., Griffith, J., & Powell, D. M. (2003). Improving decision quality: Making the case for adopting next generation site characterization practices. Remediation, 13(2), 91–111.

Crumbling, D. M., Hayworth, J. S., Call, B. A., Davis, W. M., Howe, R., Miller, D. S., et al. (2004). The maturation of the Triad approach: Avoiding misconceptions. Remediation, 14(4), 81–96.

Good, P. I., & Hardin, J. W. (2003). Common errors in statistics (and how to avoid them). Hoboken, NJ: Wiley.

Interstate Technology and Regulatory Council (ITRC). (2003, December). Technical and regulatory guidance for the Triad approach: A new paradigm for environmental project management (SCM-1). Retrieved October 1, 2004, from http://www.itrcweb.org/SCM-1.pdf

Klein, G. (1999). Sources of power: How people make decisions, Cambridge, MA: MIT Press.

Klein, G. (2003). Intuition at work: Why developing your gut instincts will make you better at what you do. New York: Doubleday.

United States Department of Energy (DOE). (2001). Adaptive sampling and analysis programs (ASAPs). DOE/EM-0592; OST/TMS ID 2946. Retrieved October 1, 2004, from http://cluin.org/download/char/ASAPs_ITSR_DOE-EM-0592.pdf

United States Environmental Protection Agency (US EPA). (2000a, January). Data quality objectives process for hazardous waste site investigations (QA/G-4HW). EPA 600/R-00/007. Retrieved October 1, 2004, from http://www.epa.gov/quality/qs-docs/g4hw-final.pdf

United States Environmental Protection Agency (US EPA). (2000b, August). Guidance for the data quality objectives process (QA/G-4). EPA 600/R-96/055. Retrieved October 1, 2004, from http://www.epa.gov/quality/qs-docs/g4-final.pdf

United States Environmental Protection Agency (US EPA). (2003, November). Guidance for obtaining representative laboratory analytical subsamples from particulate laboratory samples. EPA/600-R-03–027. Retrieved October 1, 2004, from http://clu-in.org/download/char/epa_subsampling_guidance.pdf

Deana M. Crumbling has worked in the hazardous waste site cleanup arena over the past 12 years, and has been at the US EPA since 1997. She is an analytical chemist with clinical, industrial, and research experience. She holds a BS in biochemistry, a BA in psychology, and an MS in environmental science.

Joel S. Hayworth has over 18 years of experience in the field of hazardous waste site assessment and remediation. Prior to forming Hayworth Engineering Science, Inc., in 2000, Dr. Hayworth worked at the Tennessee Valley Authority Engineering Laboratory, the Air Force Research Laboratory, and Applied Research Associates. He holds a BS in geophysics, an MS in hydrogeology, and a PhD in hydrology. He is also a professional engineer registered in the state of Florida.

Robert L. Johnson, PhD, leads Adaptive Sampling and Analysis Program (ASAP) work for the Environmental Assessment Division at Argonne National Laboratory. He has a BA in mathematics, an MS in environmental engineering, and a PhD in agricultural engineering.

Marlene Moore has over 30 years of experience working in the environmental management, sampling, and testing fields. The last ten years have been in the development and adoption of quality system management models to the environmental industry. She attended Moravian College for chemistry and various trade and business colleges. Ms. Moore is certified by the Registration Accreditation Board as an auditor for ISO 9000:2000 quality management systems and performs audits and assessments for a variety of accreditation bodies in the United States.