Triad as a Framework to Manage Decision Uncertainty

Deana M. Crumbling, M.S.
Technology Innovation Program, USEPA
Washington, D.C. (703) 603-0643
crumbling.deana@epa.gov

New Jersey DEP Consultants Day, September 4, 2003

1

Why Are We Here? NEWMOA Survey of States' Concerns

- Insufficient data to adequately determine
  - Nature & extent of contamination
  - Potential exposures
- States always want more data…recognize field analytics as mechanism to get it…BUT need more confidence in the data & staff
- Problems with the quality of project reports
  - Do not explain what was done & why
  - Need more visual aids, esp. maps

Want higher project confidence AND lower costs

2

Better Projects Are Possible!

A lot has changed since the 1980s: Experience provides

- **Better understanding of contaminated sites**

Heterogeneity Rules!!

- **R&D has produced better cleanup technologies**
  - Wider range of remedial options, **but** cost-effective selection and deployment require accurate characterization
- **R&D has produced better investigation technologies**
  - Able to capture heterogeneity
  - Produce an accurate CSM as foundation of all subsequent decisions about risk, remediation, cost-benefit assessment, insurance premiums, …

Post-1980s Tools Require Post-1980s Thinking

3

The Triad approach: A technical framework to realign cleanup program policy and engineering practice with current science & technology

4

Making Correct & Cost-Effective Project Decisions

- Correct decisions dependent on accurate understanding of site contamination
- This is termed a **Conceptual Site Model (CSM)**
- A CSM is any tool(s) that lets you represent, "conceptualize" or "model" site contamination issues and concentration populations so can make predictions about nature, extent, risk, and risk reduction strategies
- Sites often contain 2 or more discreet contaminant populations. An **accurate CSM** will separate them when **decisions** or outcomes **differ** enough to alter risks, costs, or remedial success.

5

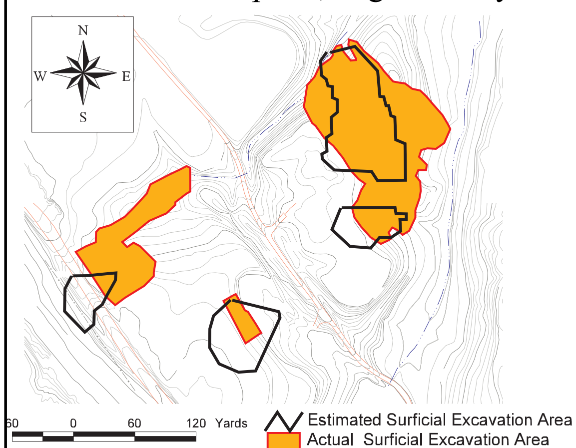
Most data uncertainty
(and thus uncertainty in the accuracy of the CSM)
stems from uncertainty about "data representativeness"

Why?

Because usually there is insufficient sampling density (samples per decision unit) to characterize a heterogeneous matrix well enough to produce an accurate CSM at scales relevant to risk and remedy decisions

6

Inaccurate Soil CSM from Traditional RI Data (black) vs. CSM from Adaptive, High Density Sampling (orange)



Excavation based on the CSM from gridded RI sampling/lab analysis would have removed ~4K c.y. clean soil and missed ~8K c.y. of contaminated soil (cf to 45K c.y. actually removed).

An accurate CSM & precision excavation saved ~$10M by getting remediation done right the first time.

Adapted from Argonne, 2002

7

GW CSM from Traditional Sampling Effort (left) vs. CSM from High Density Sampling (right)

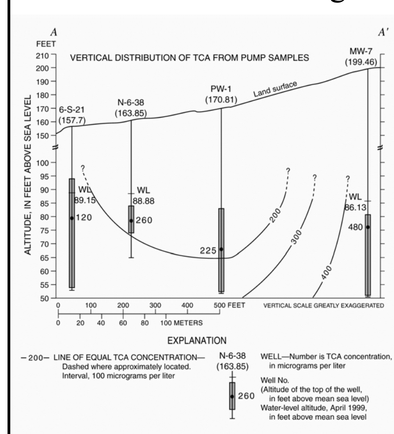


Figure 6. Vertical distribution of TCA concentrations in ground-water samples collected with the diffusion samplers and submersible pump.

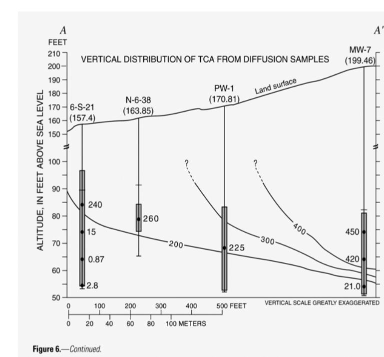


Figure 6.—Continued.

From USGS Report 02-4203 (2002). Available on-line at http://water.usgs.gov/pubs/wri/wri024203/

8

Sampling Errors Lead to Decision Errors

- **Sampling errors occur when an accurate analytical result on a tiny sample (or subsample) is erroneously assumed to represent the concentration of a larger matrix volume (or of a different matrix fraction) without controlling for sampling variables.**
- **Inaccurate CSMs and subsequent decision errors are a consequence of sampling errors.**

Triad systematic project planning helps control sampling errors by using CSMs to understand contaminant populations, and by asking "What variables could cause a decision error if uncontrolled?"

9

Triad Uses a Second-Generation Environmental Data Quality Model to Manage Data Representativeness

10

Oversimplified (First-Generation) Data Quality Model

Methods = Data = Decisions

Screening Methods → Screening Data → Uncertain Decisions

"Definitive" Methods → "Definitive" Data → Certain Decisions

This Model Fails to Distinguish: Analytical Methods from Data from Decisions

11

Data Quality Involves Much, Much More than Just Chemical Analysis

Perfect Analytical Chemistry + Non-Representative Sample(s)

↓

"BAD" DATA

Triad carefully distinguishes Analytical Quality from Data Quality

12

Any meaningful concept of data representativeness must be grounded in the <u>decision</u> context

Different decisions require different representativeness.
For example:

- A data set <u>representative of</u> a risk assessment decision usually needs to estimate the average concentration over a fairly large decision unit (called an "exposure unit")
- A data set <u>representative of</u> a cost-effective remedial design must provide information about concentration extremes and distributions at a scale specific to the remedial option considered.
- Scales for treatment technologies vary with the technology (e.g., the scale of characterization for successful chemical oxidation is much finer than for deploying *in situ* heating).

13

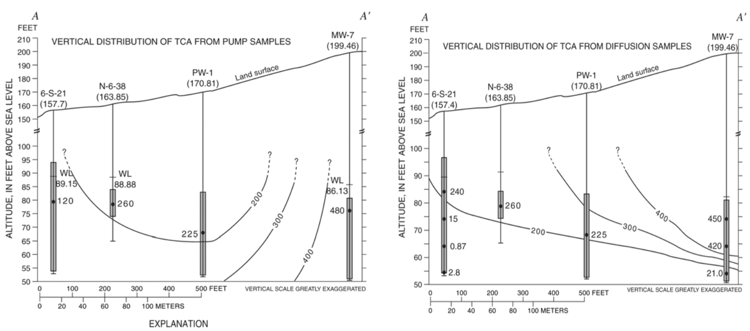
Generic Sampling Designs Cannot be Expected to Produce Representative Data for Heterogeneous Matrices

It is impossible to specify a one-size-fits-all data set that could be representative of all potential CSMs and site decisions!

Therefore, the first step of ensuring data quality is to clearly understand to what decisions the data will be applied.

14

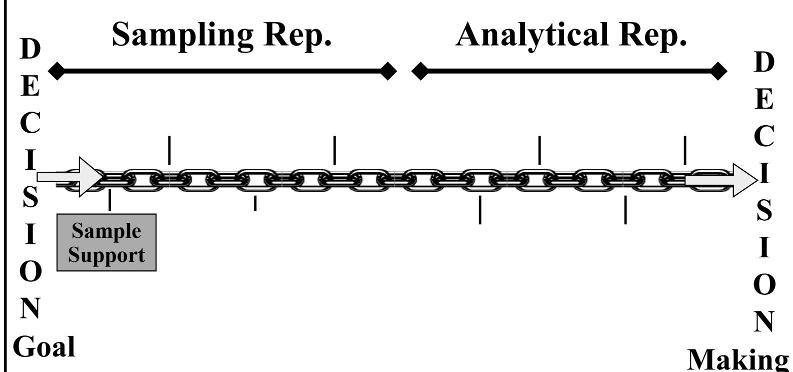
Same Analytical Method Used in Both. What Was Different?



How the sample was collected in space determined what the analytical result was!

15

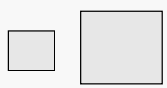
Triad Recognizes a Data Quality "Chain" for Controlling Data Generation Variables



16

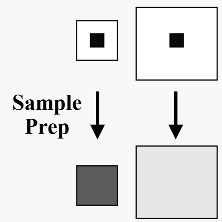
Sample Support: Includes Sample Volume

Typical regulatory and field practices assume that the size/volume of a sample has no effect on analytical results for contaminant concentrations.



That assumption doesn't hold true when environmental heterogeneity exists; **sample volume can determine the analytical result!**

The Nugget Effect

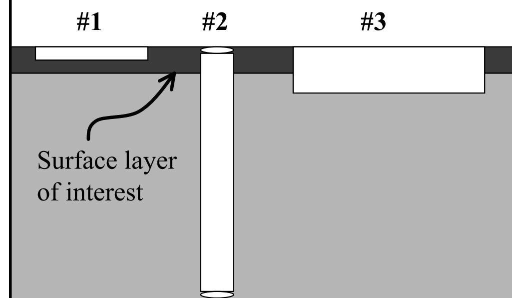


Although there is the same contaminant mass in the captured nuggets, different volumes of cleaner matrix will produce different sample concentrations after sample homogenization.

18

Sample Support: Includes Spatial Orientation

What sample support is representative of the decision?



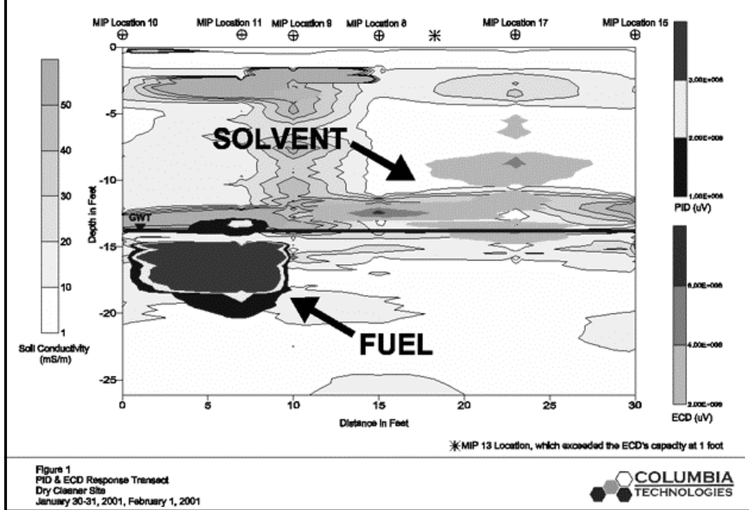
The decision driving sample collection: **Assess contamination resulting from atmospheric deposition**

Given that the dark surface layer is the soil layer impacted by atmospheric deposition relevant to this project:

Which sample support (white areas #1, #2, or #3, each homogenized before analysis) provides a sample that is representative of atmospheric deposition for this site?

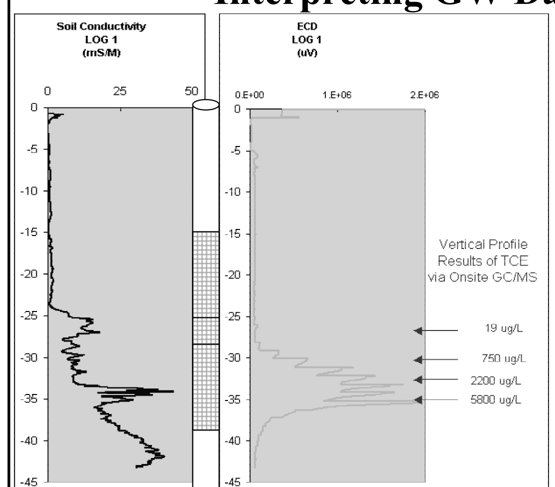
19

Control over Sample Support Critical to Distinguish Distinct Populations for Heterogeneous Contaminant Distributions



19

Sample Support Can Be Critical to Interpreting GW Data



MIP = membrane-interface probe (w/ ECD detector)

Sample support for MIP on mm-scale

Sample support for discrete field GC/MS on inch-scale

Sample support for traditional well on ft-scale

Graphic adapted from Columbia Technologies

20

Sample Support Can Spell the Difference Between Hits and NDs

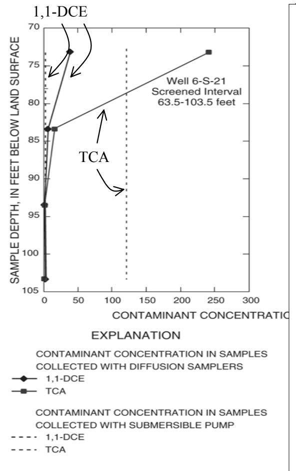


Figure 5. Comparison of selected volatile organic compound conce... and a submersible pump for wells with greater than 20-foot screened

Yet, typical regulatory and field practices are completely silent on the subject of sample support. The **assumption apparently is that sample size/volume has no effect** on analytical results for contaminant concentrations.

But this is not true.

Neglect of sample support causes it to be an uncontrolled variable producing excessive data variability, inaccurate CSMs, and conflicting data sets (for which the lab is usually blamed).

From USGS Report 02-4203 (2002); http://water.usgs.gov/pubs/wri/wri024203/

21

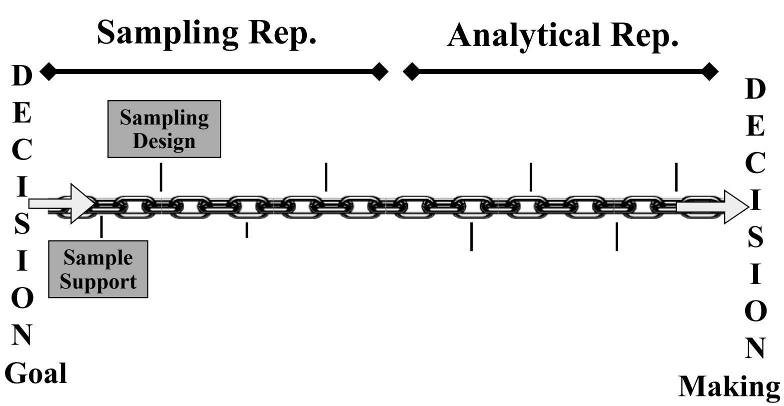
A Representative Sample Support Mimics the Decision Support

- How do you decide what sample support is correct?
- The sample support needs to be representative of the "decision support"
- Decision support = the physical properties or characteristics of the decision unit. Includes things like:
  - Matrix volume and dimensions (Ex: air deposition layer)
  - Particle sizes that are of exposure concern (Ex: fine particles with high lead concentrations that facilitate hand-mouth transmission or can be tracked or wind-blown into residences)
  - Should monitoring well data be representative of DW for exposure assessment (implies large sample support)? Or should it be representative of aquifer properties that control contaminant fate & migration or treatment success (implies small sample support)?

If decision goals are unknown, then decision support unknown! Then is *impossible to plan for representative data collection*!!!

22

More Variables in the Data Quality "Chain"



23

Can Your Sampling Design Avoid Decision Errors from Misleading Grab Sampling?

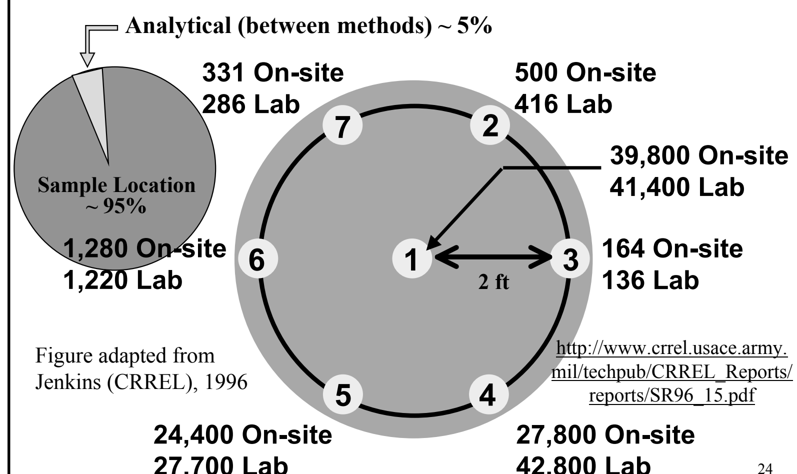
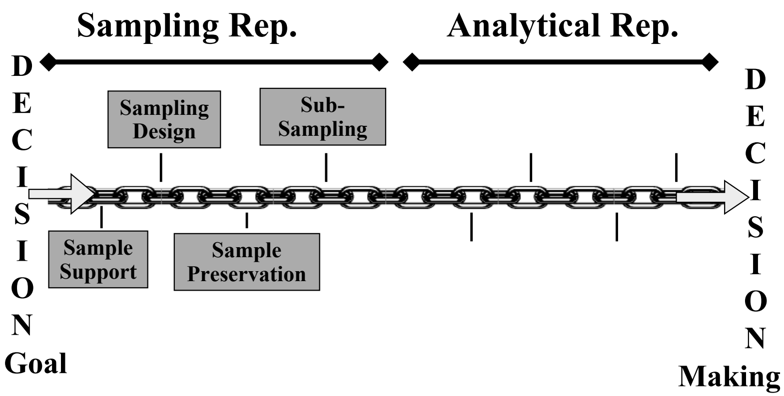


Figure adapted from Jenkins (CRREL), 1996

http://www.crrel.usace.army.mil/techpub/CRREL_Reports/reports/SR96_15.pdf

24

More Variables in the Data Quality "Chain"



All links in the Data Quality chain must be intact for data to be representative of the decision!

25

Is the Subsample Support Representative?

Lead Concentration Varies w/ Particle Size

Soil Grain Size (Standard Sieve Mesh Size)	Soil Fraction-ization (%)	Pb Conc. in fraction by AA (mg/kg)	Lead Distribution (% of total lead)
Greater than 3/8" (0.375")	18.85	10	0.20
Between 4-mesh and 3/8"	4.53	50	0.24
Between 4- and 10-mesh	3.65	108	0.43
Between 10- and 50-mesh	11.25	165	2.00
Between 50- and 200-mesh	27.80	836	25.06
Less than 200-mesh	33.92	1,970	72.07
Totals	100%	927 (wt-averaged)	100%

Adapted from ITRC (2003); see http://www.itrcweb.org/SMART-1.pdf

Lab subsampling that captures larger particles will get lower results!!

26

Is the Subsample Support Representative?

^{241}Am Concentration Varies w/ Subsample Support

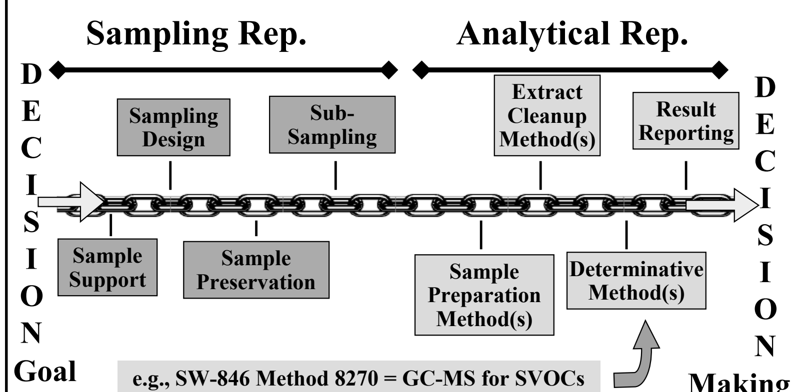
Subsample Support (after sample was dried, ball-milled, sieved <10-mesh)	Coefficient of Variation	Number of subsamples required to estimate the sample true mean ± 25% *	Number of subsamples required to estimate the sample true mean ± 10% *
1 g	0.79	39	240
10 g	0.27	5	28
25 g	0.30	6	35
50 g	0.12	1	6
100 g	0.09	1	4

* Using classical parametric statistics at 95% confidence   Adapted from DOE (1978)

Major problem!! Advancing analytical technologies use smaller and smaller subsample aliquots--undermine representativeness!

27

The Data Quality "Chain"



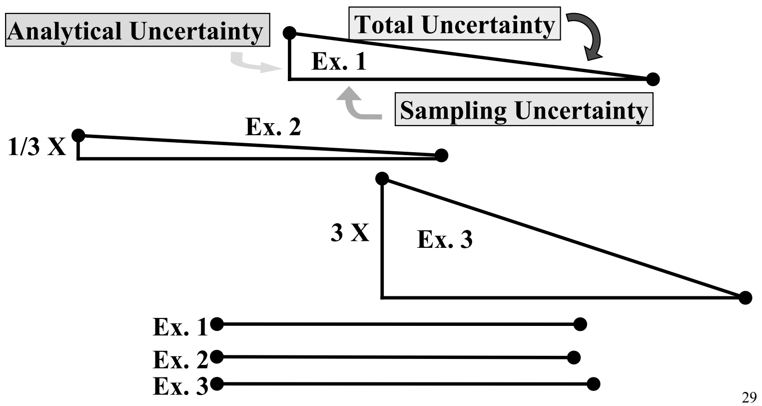
All links in the Data Quality chain must be intact for data to be representative of the decision!

28

Summing Uncertainties

Uncertainties add according to ($a^2 + b^2 = c^2$)

Decreasing or increasing analytical uncertainty has little effect on overall data quality if sampling uncertainty is unchanged



29

Sampling vs. Analytical Variability

Example using a Brownfields project data set (scrap yard site with contaminated soil)

Std Dev $_{Sampling}$: Std Dev $_{Analytical}$ = Samp:Anal Ratio

(Total variability determined from entire data set. LCS data used to estimate analytical variability. Sampling variability calculated by subtraction.)

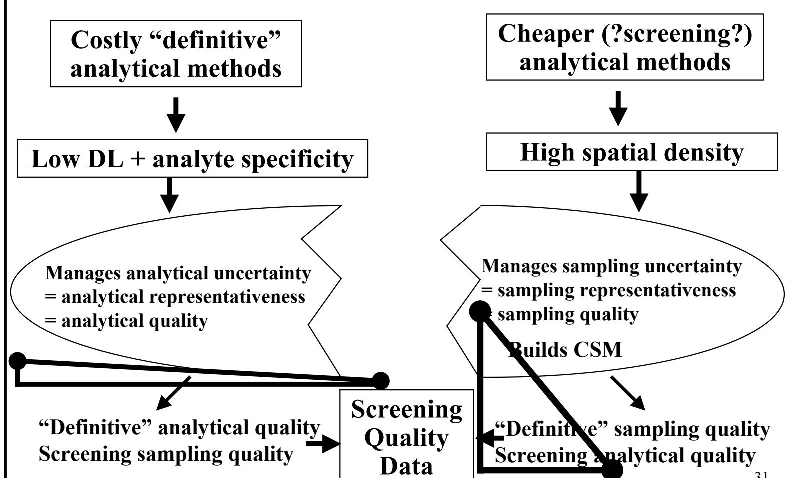
As (natural background present): 22.4 : 7 = 3 : 1
Pb (high spatial variability): 3255 : 3 = 1085 : 1

A 3:1 ratio for sampling-to-analytical Std Dev = 90% of statistical variance due to non-method considerations

A 1000:1 ratio for sampling-to-analytical Std Dev = 99.999% of statistical variance due to non-method considerations

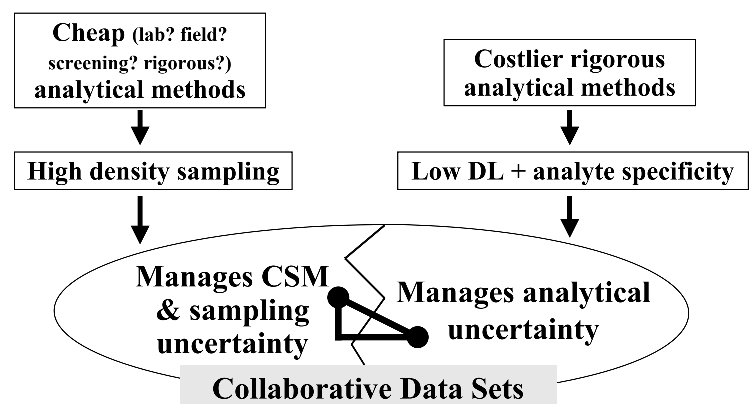
30

The Strengths & Limitations of Methods



31

Data Quality for Heterogeneous Matrices



Collaborative data sets complement each other so that all sources of data uncertainty important to the decision are managed

32

Data Quality for Heterogeneous Matrices

Cheap (screening?) analytical methods → **High spatial density** → **Manages CSM & sampling uncertainty**

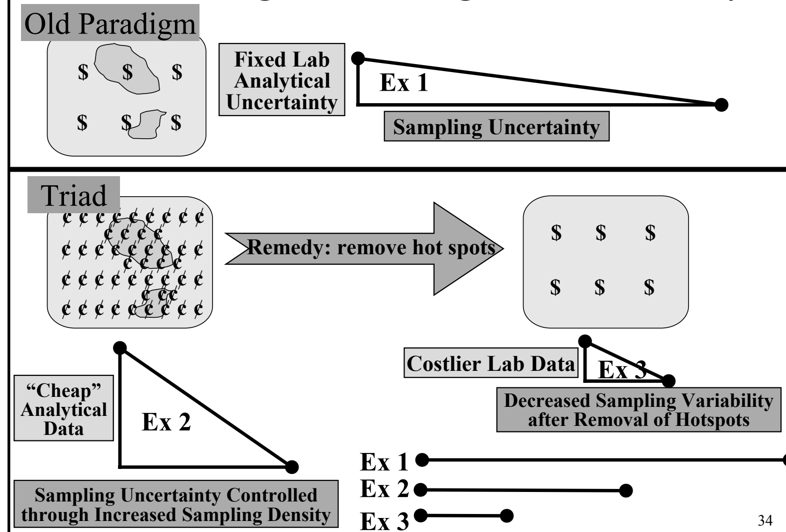
Costly definitive analytical methods → **Low DL + analyte specificity** → **Manages analytical uncertainty**

Collaborative Data Sets

Collaborative data sets complement each other so that all sources of data uncertainty important to the decision are managed

33

A New Paradigm that Manages Decision Quality



34

Integrating these Concepts into Practice
The Triad Approach

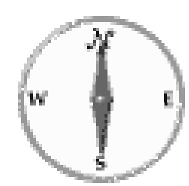
Systematic Project Planning

Dynamic Work Plan Strategy

Real-time Measurement Technologies

35

The Triad approach uses the concept of "managing uncertainty" as a compass that charts a clear course through the complexities of site cleanup science and policy.



36

The Triad is NOT…

- …written in all caps (not an acronym!)
- …just about using field analytical! (Warning: Just using field analysis does not mean they used the Triad approach!!)
- …a way to justify using field analysis without using proper QC.
- …a license to escape regulatory oversight or accountability by writing vague work plans.

37

The Triad IS about…

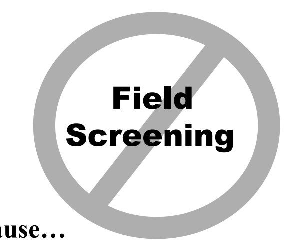
- Improving project quality by actively managing DECISION uncertainty using new tools & strategies
- Constructing accurate CSMs (as a primary Triad product!) to support cost-effective decisions
  - Using real-time decision-making to cut lifecycle costs
  - Updating the environmental data quality model to include sampling variables & tailoring QC to manage uncertainty
- Managing uncertainty in communications with solid “connect-the-dots” documentation and unambiguous terminology
- Cultivating professional competence & multidisciplinary teams (“allied environmental professionals”)

38

Triad relies on internally consistent and intuitive terminology to use decision uncertainty management as an anchor for data quality concepts

39

Misleading Terminology



Misleading because…

- Not all methods run in the field are screening methods!
- Not all data produced in the field are screening quality data!
- Fixed labs using definitive analytical methods may produce screening quality data!
- Screening methods can (and should) be used more often in fixed labs to better manage sampling uncertainty and improve analytical performance of traditional methods.

40

The Diffusion of Innovation

"At first people refuse to believe that a strange new thing can be done, then they begin to hope it can be done—then it is done and all the world wonders why it was not done centuries ago."

—Francis Hodges Burnett

41

On-Line Resources

- EPA TIO's Clu-In website http://cluin.org/triad
- Field Analytics Encyclopedia website http://fate.cluin.org
- ITRC website http://www.itrcweb.org
- Argonne ASAP website http://www.ead.anl.gov/project/dsp_topicdetail.cfm?topicid=23
- DOE DQO/statistics training materials website http://www.hanford.gov/dqo/training/contents1.html
- USACE Engineering Manuals (EMs) http://www.usace.army.mil/inet/usace-docs/eng-manuals/em.htm
- US Army Corps CSM Guidance EM: EM 1110-1-1200
- USACE CRREL Technical Report Library http://www.crrel.usace.army.mil/products/products.html

42

Assorted Papers/Articles

- EPA TIO Triad information: http://cluin.org/triad
  - *ES&T* "Managing Uncertainty in Environmental Decisions" article: http://cluin.org/download/char/oct01est.pdf
  - *Quality Assurance* journal "Representativeness" article: http://cluin.org/download/char/dataquality/dcrumbling.pdf
  - *Remediation* journal "Next Generation Practices" article: http://cluin.org/download/char/spring2003v13n2p91.pdf
- Assorted topical papers on environmental data quality concerns: http://cluin.org/products/dataquality/
- Assorted EPA guidance on representative sampling: http://cluin.org/char1_edu.cfm#samp_coll and http://cluin.org/char1_edu.cfm#stat_samp

43