Real-Time Data Collection And Interpretation “Why Is It Important?”

Stephen Dyment
US EPA
Technology Innovation Field Services Division
dyment.stephen@epa.gov
First A Little History

“Doubt is not a pleasant condition, but certainty is absurd.”

Voltaire, humanist, rationalist, & satirist (1694 -1778)

“As we know, there are known knowns. There are things we know we know. We also know there are known unknowns. That is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know.”

Donald Rumsfeld, Feb. 12, 2002
Department of Defense
Decision Quality Only as Good as the Weakest Link in the Data Quality Chain

Each link represents a variable contributing toward the quality of the analytical result. All links in the data quality chain must be intact for data to be of decision-making quality!
Sampling vs. Analytical Uncertainty

Analytical = 5%
Sampling = 95%

331 On-site 286 Lab
500 On-site 416 Lab
39,800 On-site 41,400 Lab
164 On-site 136 Lab
24,400 On-site 27,700 Lab
27,800 On-site 42,800 Lab

Figure adapted from Jenkins (CRREL), 1996
Uncertainty Math Magnifies Weakest Link’s Effects in Data Quality Chain

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

**Example:**
- AU = 10 ppm, SU = 80 ppm: \(TU = 81\) ppm
- AU = 5 ppm, SU = 80 ppm: \(TU = 80\) ppm
- AU = 10 ppm, SU = 40 ppm: \(TU = 41\) ppm
- AU = 20 ppm, SU = 40 ppm: \(TU = 45\) ppm
Real-Time Data

- Really “Within-Time”
  - Results/information in time to facilitate dynamic work strategies
  - Field analytical methods (XRF, IA, UV, colorimetric)
  - Direct sensing tools (MIP, LIF, EC, CPT)
  - Geophysical tools (EM, resistivity, GPR)
  - Mobile labs, field GC, modified methods
  - Quick turn fixed based lab analysis
- Can benefit from “demonstration of method applicability” (DMA), field based investigation levels, decision logic, decision support tools
"I think that in the discussion of natural problems we ought to begin not with the Scriptures, but with experiments, and demonstrations.”

**Galileo Galilei**

- Concept founded in SW-846, performance based measurement (PBMS) initiative
  [http://www.epa.gov/sw-846/pbms.htm](http://www.epa.gov/sw-846/pbms.htm)

- Initial site-specific performance evaluation
  - Analytical and direct sensing methods
  - Sample design, sample collection techniques, sample preparation strategies
  - Used to select information sources for field and off-site

- Goal is to establish that proposed technologies and strategies can provide information appropriate to meet project decision criteria
Just A Few DMA Benefits

- Augment planned data collection and CSM development
- Test drive Decision Support Tools (DSTs)
  - Sampling and statistical tools
  - Visualization tools, data management tools
- Develop relationships between visual observations and direct sensing tools
- Flexibility to change tactics based on DMA rather than full implementation
- Establish initial decision logic for DWS
- Evaluate existing contract mechanisms
- Optimize sequencing, load balance, unitizing costs
- The “Brownfields perception”
Example Correlations Between LIF Response and Free Product

Free Product At >50% Relative Fluorescence for Gasoline

Presence of free product unlikely

Presence of free product likely

Free Product At >75% Relative Fluorescence for Oil

Presence of free product unlikely

Presence of free product likely
Real-Time Data Allows. . .

- The use of dynamic work strategies
- Continuous updates to CSM
- Target and manage greatest uncertainties as they are identified
- Optimize well placement/screen interval
- Limit mobilizations/prioritize data gaps
- Choose collaborative samples
- Combine traditionally segmented activities/
  Monitor in-situ remedy effectiveness
Collaborative Data Sets Address Analytical and Sampling Uncertainties

- **Cheaper/rapid** (lab? field? std? non-std?) analytical methods
  - Targeted high density sampling

- **Costlier/rigorous** (lab? field? std? non-std?) analytical methods
  - Low DL + analyte specificity

Manages CSM & sampling uncertainty

Manages analytical uncertainty

Collaborative Data Sets
Real-time Measurement Systems and Collaborative Data Sets

- Field based action levels or investigation levels
  - Levels for field analytics or direct sensing tools that trigger action
    - Collection of collaborative data
    - Step outs, additional sampling or analysis, well placement, etc.
  - Remedy implementation
    - Removal
    - Confirmation of clean (sometimes required)

*DQOs for Superfund* guidance
“For the data to be definitive, either analytical or total measurement error must be determined.” (p. 43)
Lead Niton vs. ICP

59 Total pairs

\[ y = 1.0222x + 34.612 \]
\[ R^2 = 0.946 \]

10 False Positive Errors = 26%

True Positive 20 Pairs

0 False Negative Error = 0%

True Negative 29 Pairs
3 Way Decision Structure With Region of Uncertainty

Lead Niton vs. ICP

\[ y = 1.0222x + 34.612 \]

\[ R^2 = 0.946 \]

- 3 False Positive Errors = 7.7%
- True Positive Pairs = 19
- 11 Samples for ICP
- Too Close to Call
- 0 False Negative Error = 0%
- True Negative Pairs = 26
- Upper Field Investigation Level = 450 ppm
- Lower Field Investigation Level = 350 ppm
Real-Time Data Interpretation

- Critical Part II of Real-Time
  
  *Collect, process, visualize, make decisions, communicate with stakeholders*

- Visualization software
  - Freeware like SADA, FIELDS, RAT, VSP
  - Commercial products: EVS, GMS, EQuIS

- Decision support tools
  - Commercial and proprietary software
Dynamic Work Strategies

- From this...
- To this......
- Detailed decision logic
- Program
- Project
- Field effort
- Sampling
- Contaminant

We took a few samples... What do we do next?

We should stop, write a report, identify data gaps, and write another workplan!
Conceptual Site Models

Figure 4-5: Conceptual Model of Exposure Pathways

Notes:
- X = Pathway broken at this point
- Shaded boxes indicate exposure pathway is not complete at this point

Available information indicates that predominant land use is residential. Residential assumptions will provide the highest frequency and duration of exposure. Potential exposures to notification workers could occur through emissions during disturbance of areas where VOCs have been sampled.

Potential exposure to notification workers could occur from VOCs in groundwater that is assumed to be used as a domestic or industrial water supply based on poor water quality and limited yield.
Decision Support Tools

- From this...
- To this......

Geostatistics using SADA

Evergreen Berm, Plan View
Probability that 1-ft Deep Volumes > 250 ppm Pb
Statistical Sampling Design

Choose wisely: It Makes a Big Difference in Sample Numbers!

Actual mean closer to AL --->

<table>
<thead>
<tr>
<th>GR width: StDev</th>
<th>350 ppm</th>
<th>200 ppm</th>
<th>100 ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ppm</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>200 ppm</td>
<td>5</td>
<td>10</td>
<td>36</td>
</tr>
<tr>
<td>300 ppm</td>
<td>8</td>
<td>21</td>
<td>79</td>
</tr>
</tbody>
</table>

Increasing variability --->

Achilles and Innovative Strategies
Sampling Design

- Triangular grid program
- 203 samples allocated
- Observed error rates:
  - Missed contamination: 0 ft²
  - Incorrectly excavated clean: 3,500 ft² (35% over-excavation)
Initial Conceptual Site Model

- Based on soft information, assign probability of contamination being present
- Map shows this CSM pictorially, color-coded based on contamination probability
- This CSM drives subsequent sampling decisions & becomes an important point of concurrence for stakeholders
GeoBayesian Adaptive Sampling Progression
Adaptive Program Performance

- Completely done with 62 samples
- After only 22 samples, outperformed traditional 203 sample grid program from an error rate perspective
- Requires real-time measurements
Another Example....

- 4 residential backyards screened by XRF for arsenic with action level of 25 ppm averaged over yard
- Use XRF to determine:
  - whether each yard is likely above or below action level, and
  - if below, how many laboratory samples are required to statistically show it?
Here's what the yards look like:

- Yard 1: 2 samples
- Yard 2: 2 samples
- Yard 3: 2 samples
- Yard 4: 6 samples

**Arsenic (ppm):**
- 4 - 9
- 10 - 15
- 16 - 25
- 26 - 50
- 51 - 270

**Test Parameters:**
- Action level = 25 ppm
- False negative error rate = 0.05
- False positive error rate = 0.05

**How many samples are required?**
- Yard 1: average = 24 ppm, stdev = 41 ppm
- Yard 2: average = 6.8 ppm, stdev = 0.7 ppm
- Yard 3: average = 7.0 ppm, stdev = 0.7 ppm
- Yard 4: average = 10 ppm, stdev = 9 ppm
Real-Time Data Experiences
Lessons Learned

- Linear regression - can be helpful or misleading
- Heterogeneity - large scale, small scale, and within sample
  - Don’t expect collaborative data to compare any better than 2 labs or even the same lab
- Focus on decision quality
- Data management and communication are key
- Structure vendor contracts to include some DMA principles
- Evaluate contingencies
- Particular instruments ≠ technology generalizations
Brownfields Technical Support Center
http://www.brownfieldstsc.org/index.cfm

- Technical support services
  - Facilitate systematic planning
  - Build/refine conceptual site model
  - Sampling design, scope of work, sequencing
  - Real-time measurements, field analytics/collaborative data sets
  - Statistical data analysis
  - Decision support tools
  - Work-plan development/review
  - Remedy optimization
  - Monitoring network optimization
  - Independent design reviews
Comprehensive Sources

- Hazardous Waste Cleanup Information (CLU-IN) Internet Site
  - http://clu-in.org
- Triad Resource Center
  - http://www.triadcentral.org
- EPA Internet site
  - http://www.epa.gov
- Federal Remediation Technologies Roundtable
  - http://www.frtr.org
- Field Analytical Technology Encyclopedia (FATE)
  - http://fate.clu-in.org/
- Brownfields Technology Support Center
  - http://www.btsc.org
- Interstate Technology Regulatory Council (ITRC)
  - http://www.itrcweb.org