Data Quality:
Closing the Decision-Data Loop

Deana Crumbling
USEPA Technology Innovation Office
crumbling.deana@epa.gov
USACE Chemists' Business Meeting
March 6, 2002

What Is the Vision?

- **Triad Approach** = Integration of **systematic planning**, **dynamic work plans**, and **real-time analysis** as applied to wastes and contaminated sites

  ⇓ time & costs; ⇑ decision certainty

- Theme for the Triad Approach = Managing the largest sources of decision error, especially the sampling representativeness of data

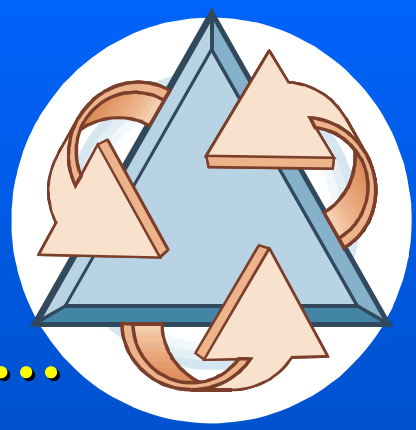
A Systems Approach Framework
The Triad Approach

Systematic Planning

Dynamic Work Plans

Real-time Measurement Technologies

Unifying Concept for Triad: Managing Uncertainty



Systematic planning is used to proactively…

- **Manage uncertainty about project goals**
  - Identify decision goals with tolerable overall uncertainty
  - Identify major uncertainties (cause decision error)
  - Identify the strategies to manage each major uncertainty
- **Manage uncertainty in data**
  - **Sampling uncertainty:** manage sample representativeness
  - **Analytical uncertainty:** especially if field methods are used
- **Draw upon multidisciplinary expertise**
  - **A TEAM** is the best way to bring needed knowledge to bear

Dynamic Work Plans

- Real-time decision-making "in the field"
  - Evolve CSM in real-time
  - Implement pre-approved decision tree using senior staff
  - Contingency planning: most seamless activity flow possible to reach project goals in fewest mobilizations
- Real-time decisions need real-time data
  - Use off-site lab w/ short turnaround?
  - Use on-site analysis?
    - » Use mobile lab with conventional equipment?
    - » Use portable kits & instruments?

Mix And Match

In all cases, must generate data of <u>known</u> quality

Generating Real-time Data Using Field Methods

Manage Uncertainty through Systematic Planning

- **Need clearly defined data uses—tie to project goals**
- **Understand dynamic work plan—branch points & work flow**
- **Project-specific QA/QC protocols matched to intended data use**
- **Select field analytical technologies to**
  - Support the dynamic work plan (greatest source of $$ savings)
  - Manage sampling uncertainty (improves decision quality)
- **Select fixed lab methods (as needed) to**
  - Manage uncertainties in field data (as ONE aspect of QC)
  - Supply analyte-specific data and/or lower quantitation limits (as needed for regulatory compliance, risk assessment, etc.)

Updating the Data Quality Concept as a Tool to Achieve Decision Quality

Data is Generated on Samples

Perfect Analytical Chemistry + **Non-Representative Sample**

"BAD" DATA

Distinguish:
Analytical Quality from Data Quality

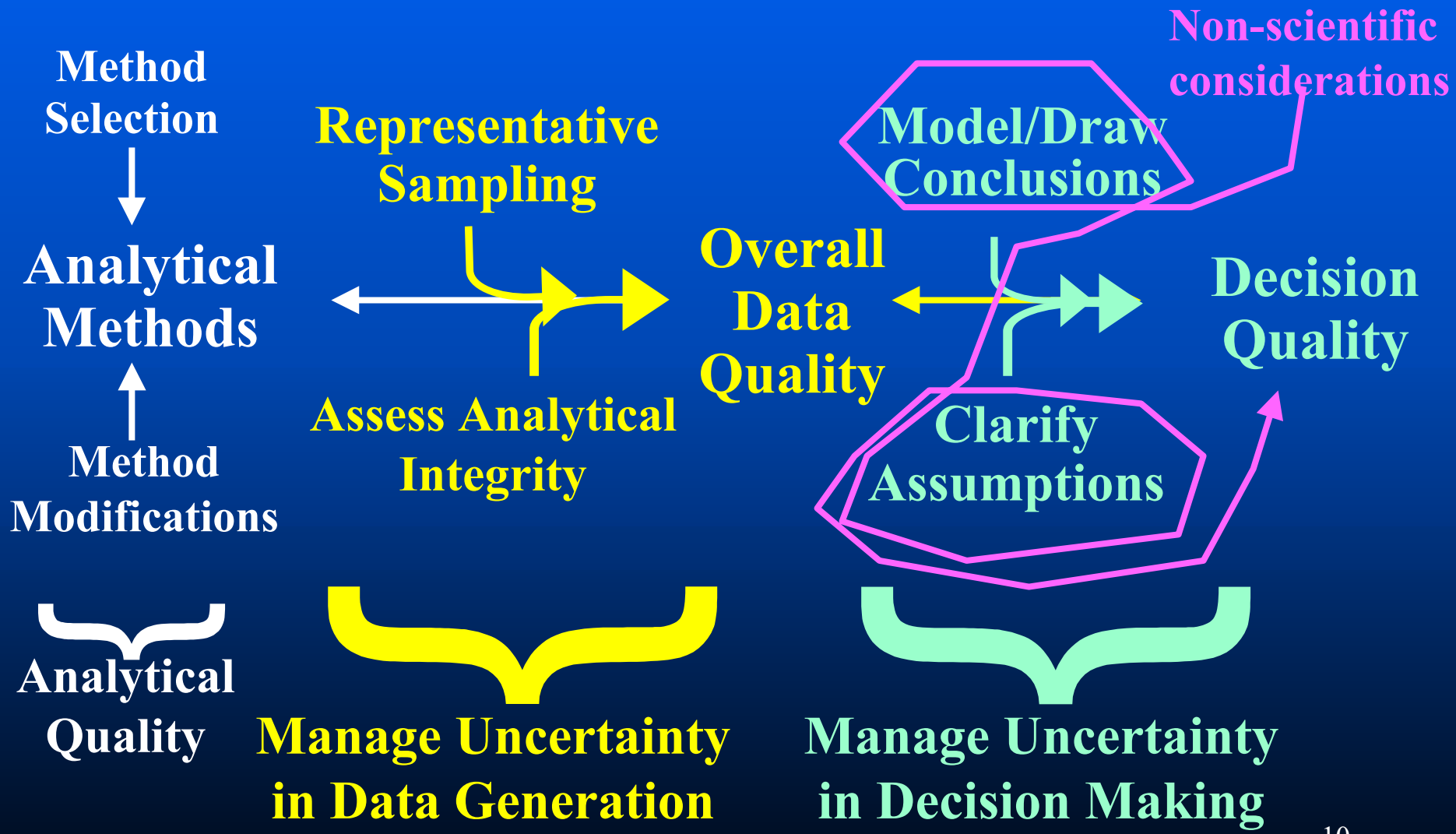
Oversimplified Data Quality Model

Methods = Data = Decisions

Screening Methods →	Screening Data →	Uncertain Decisions
"Definitive" Methods →	"Definitive" Data →	Certain Decisions

Distinguish:
Analytical Methods from Data from Decisions

Distinguishing Concepts



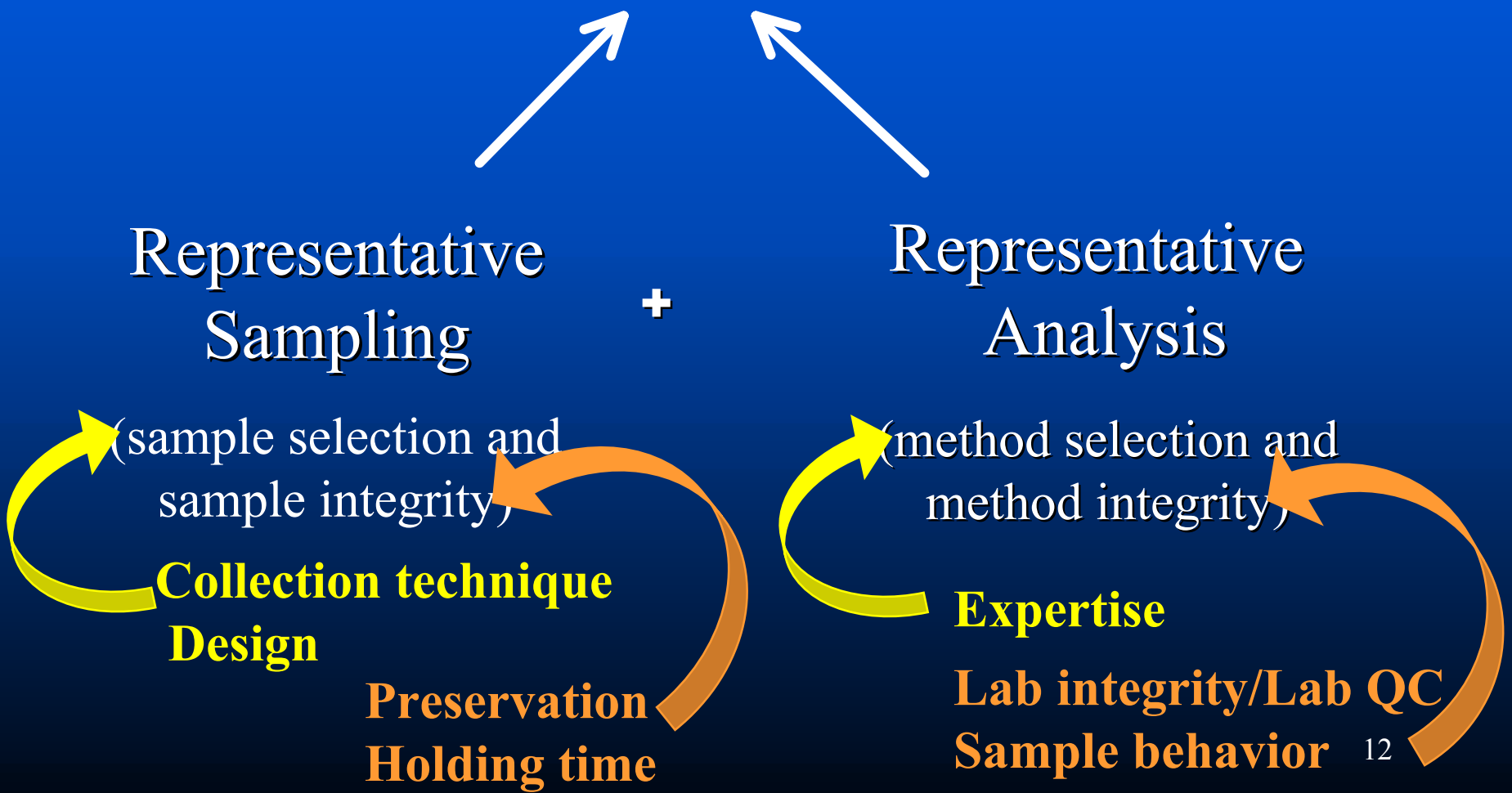
What is "Data Quality"?

Data Quality = The ability of data to provide information that meets user needs

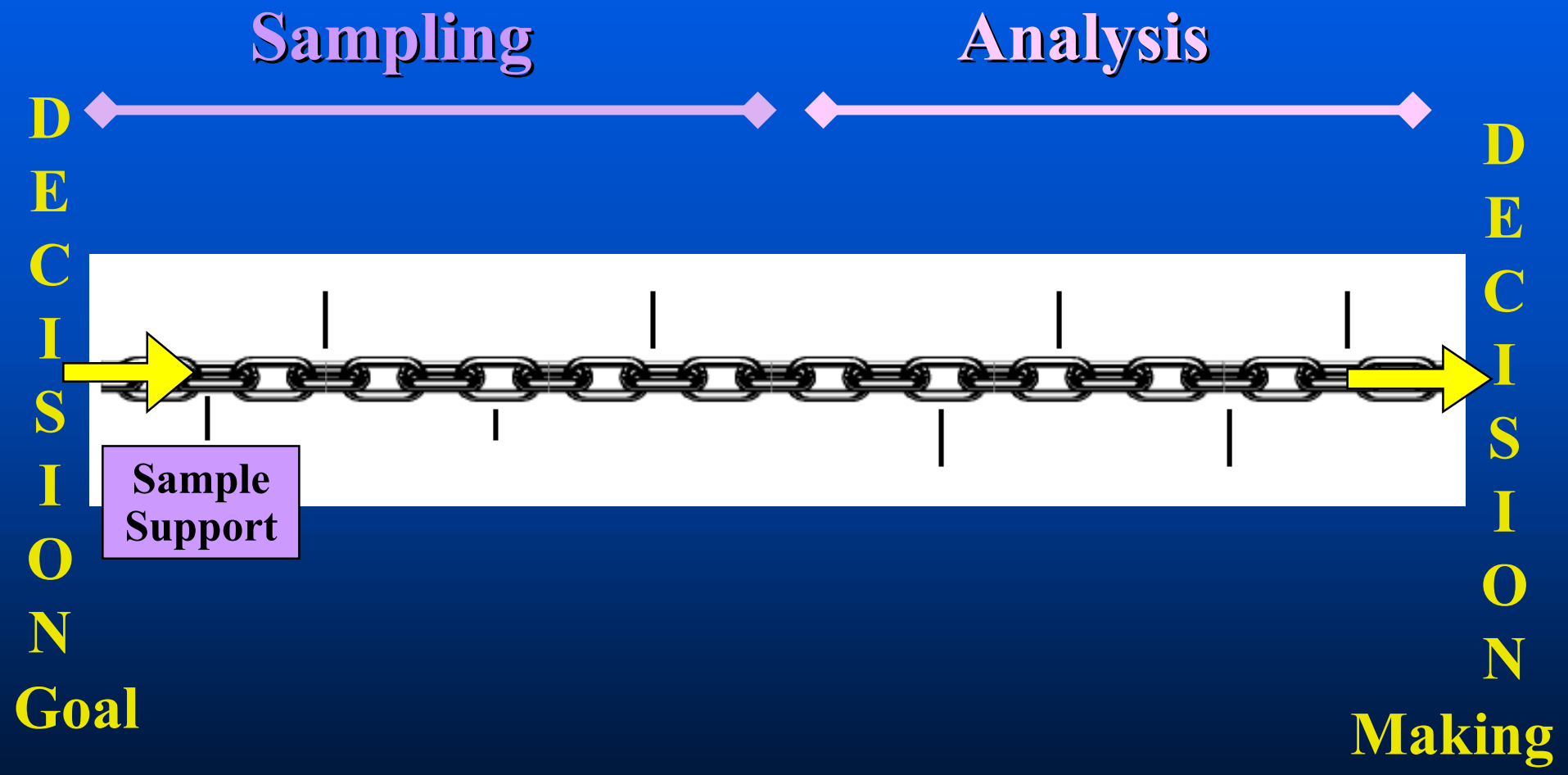
- Users need to make correct decisions
- Data quality is a function of data's…
  - ability to represent the "true state" in the context of the decision to be made
    - » The decision defines the scale for representing the "true state"
  - information content (including its uncertainty)

Overall Data Quality

Proper Interpretation

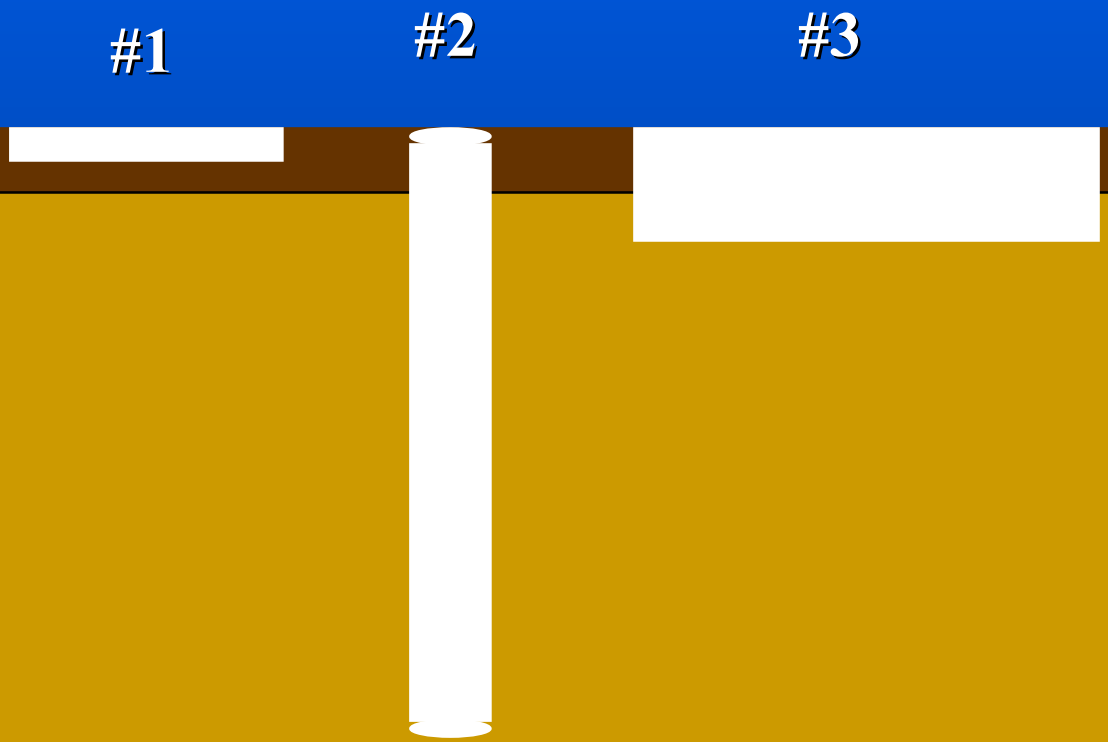


The Data Quality "Chain"



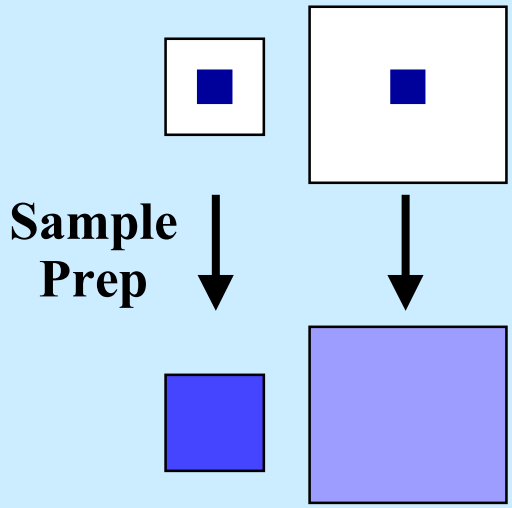
Sample Support: Critical to Representativeness

Sample Volume & Orientation



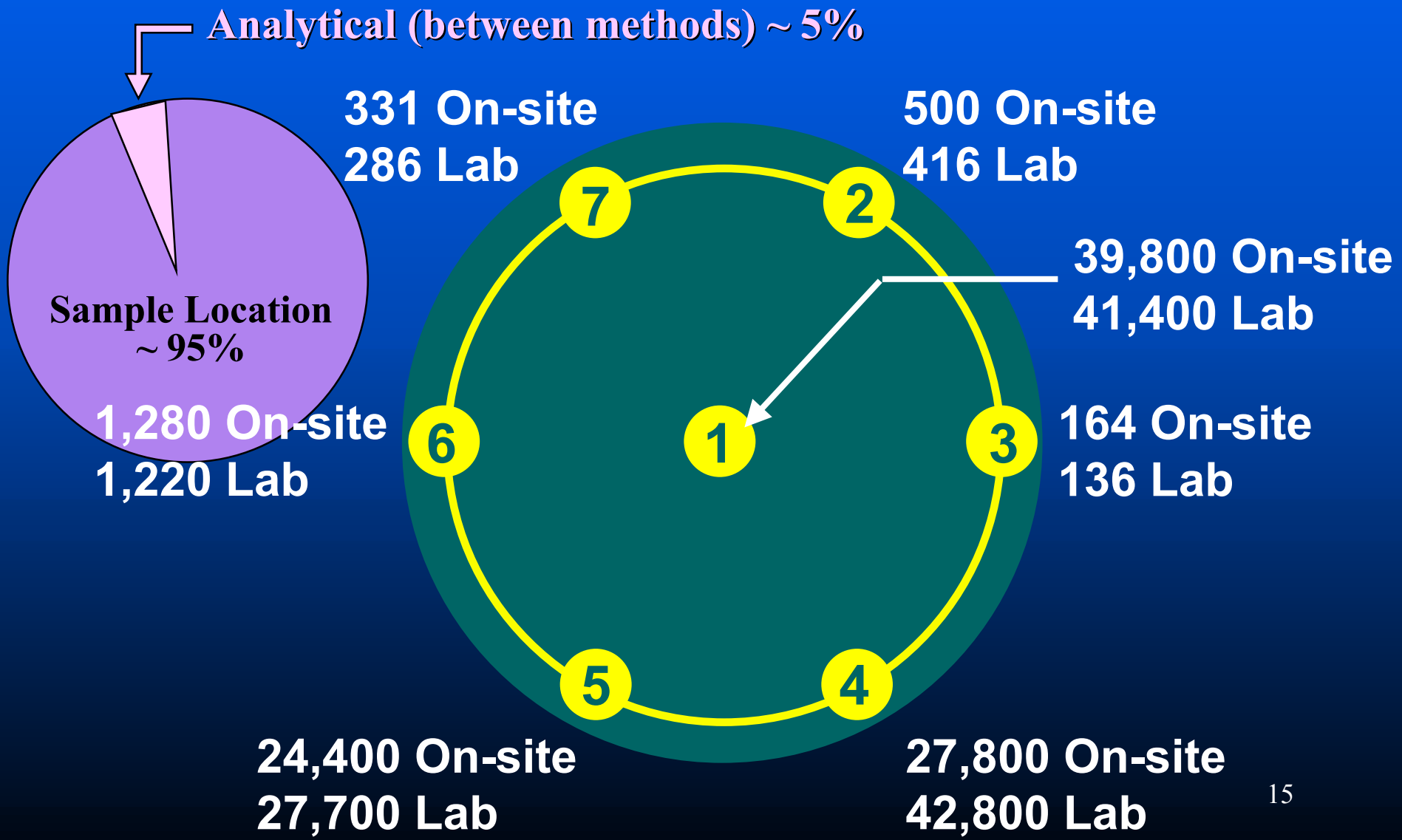
The decision driving sample collection: Assess contamination resulting from atmospheric deposition

The Nugget Effect

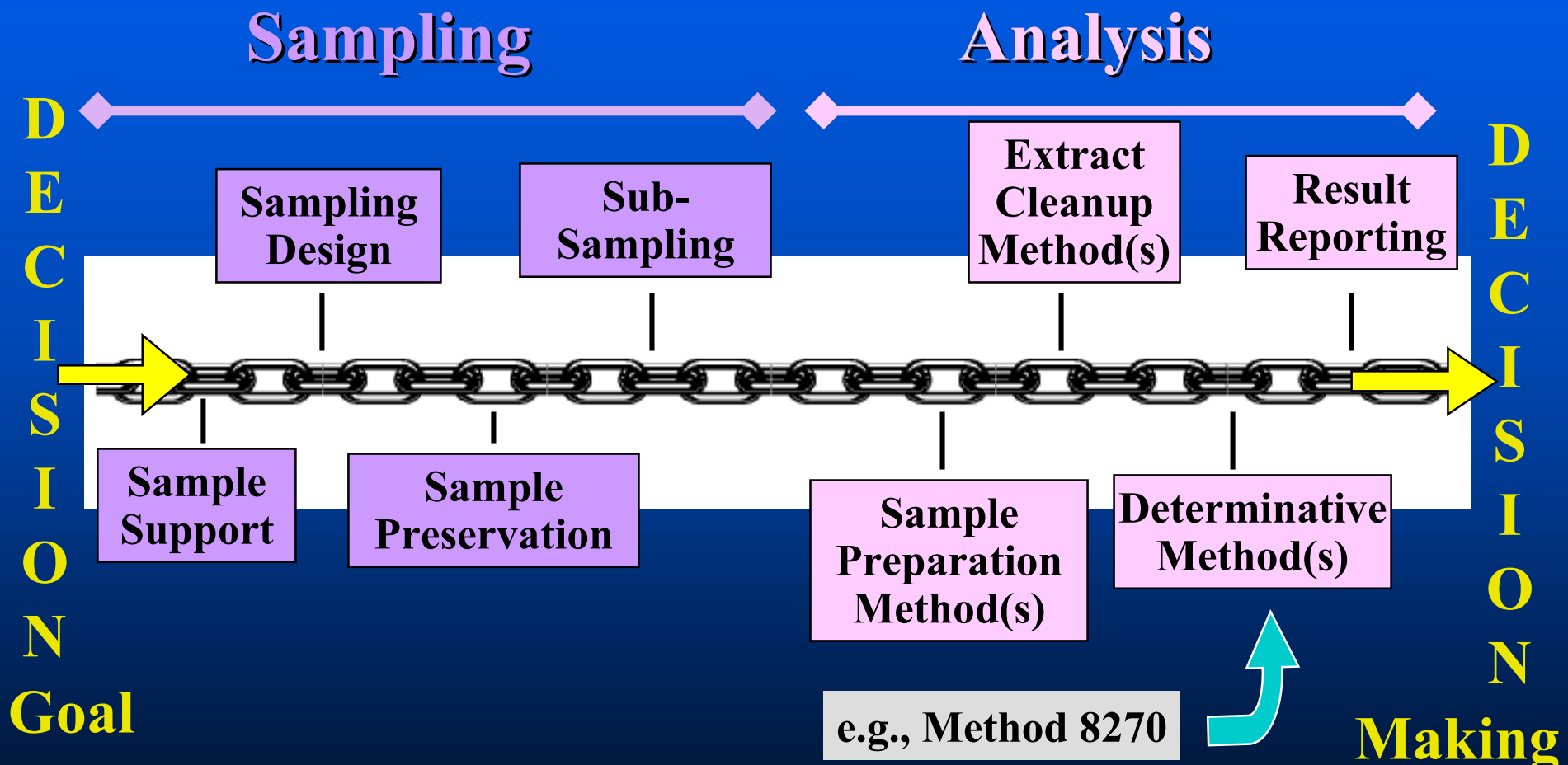


Same Contaminant Mass in Nugget, but Different Sample Volumes Produce Different Concentrations

Partitioning Variability: Sample Location vs. Analytical Method



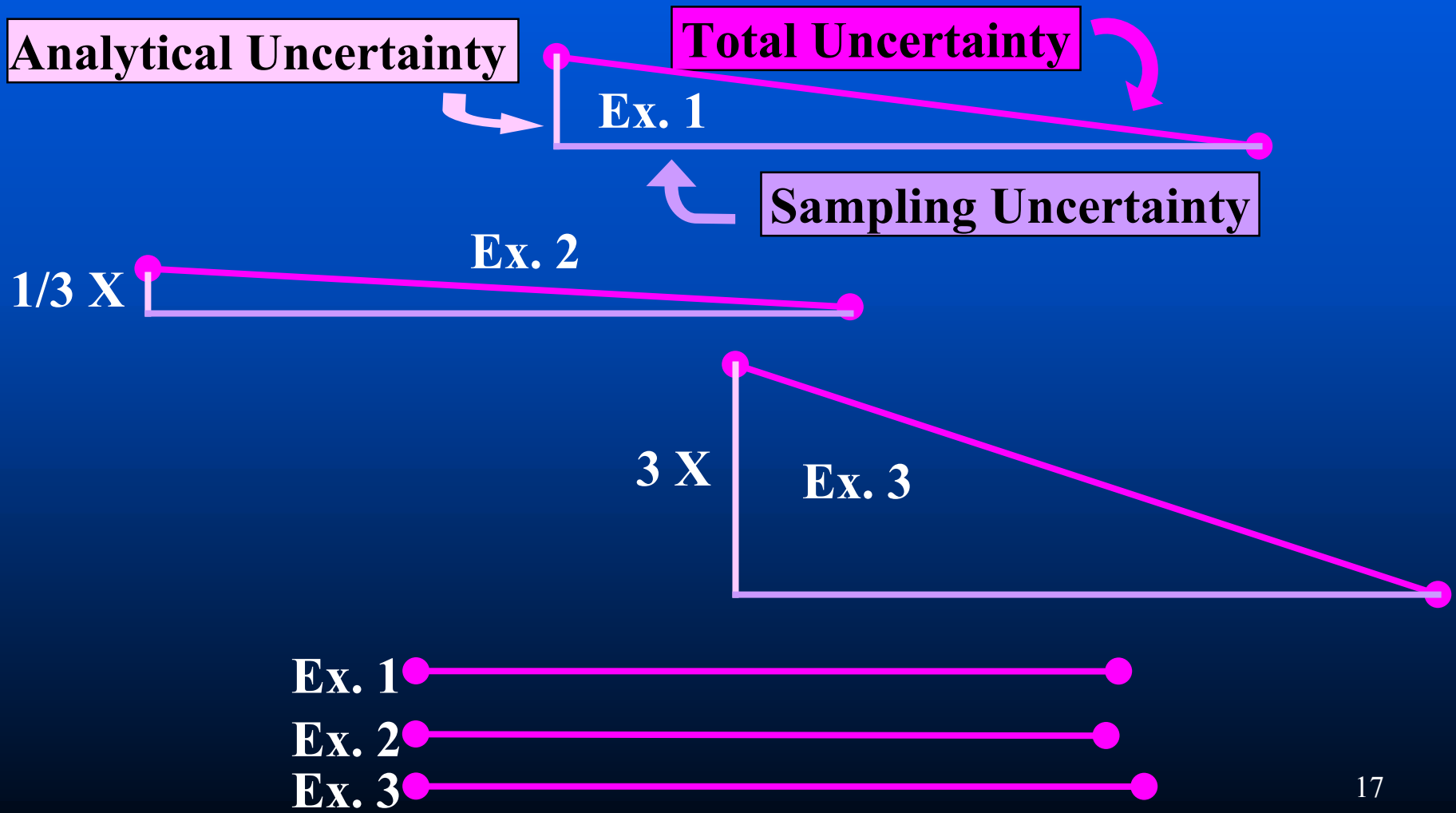
The Data Quality "Chain"



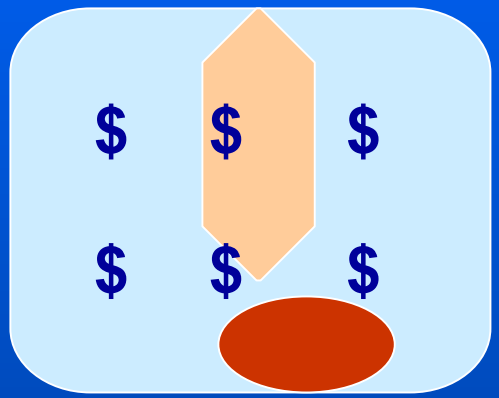
All links in the Data Quality chain must be intact for Decision Quality to be supported !

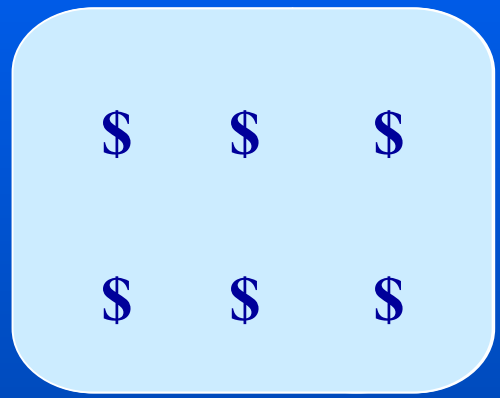
Summing Uncertainties

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



Decision Quality vs. Analytical Quality





Fewer higher analytical quality data **points** ⇒ **Lower** information value of the data **set**

NOW, few higher analytical quality data **points** ⇒ **Highly** informative data **set**

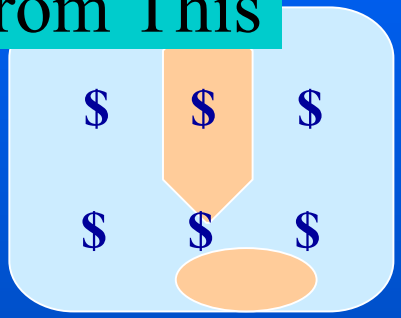
Less likely

Nearly Certain

Goal: A defensible site decision that reflects the "true" site condition

Improve Decision Quality--Manage Uncertainties

From This

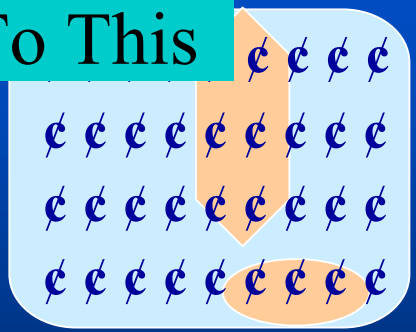


Fixed Lab Analytical Uncertainty

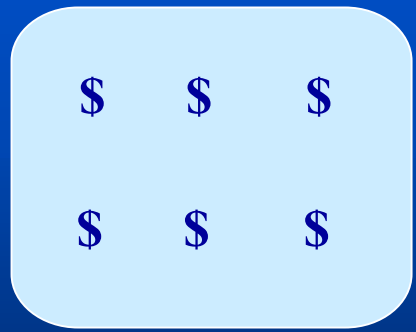
Ex 1

Sampling Uncertainty

To This



Remove hot spots



Field Analytical Data

Ex 2

Sampling Uncertainty Controlled through Increased Density

Fixed Lab Data

Ex 3

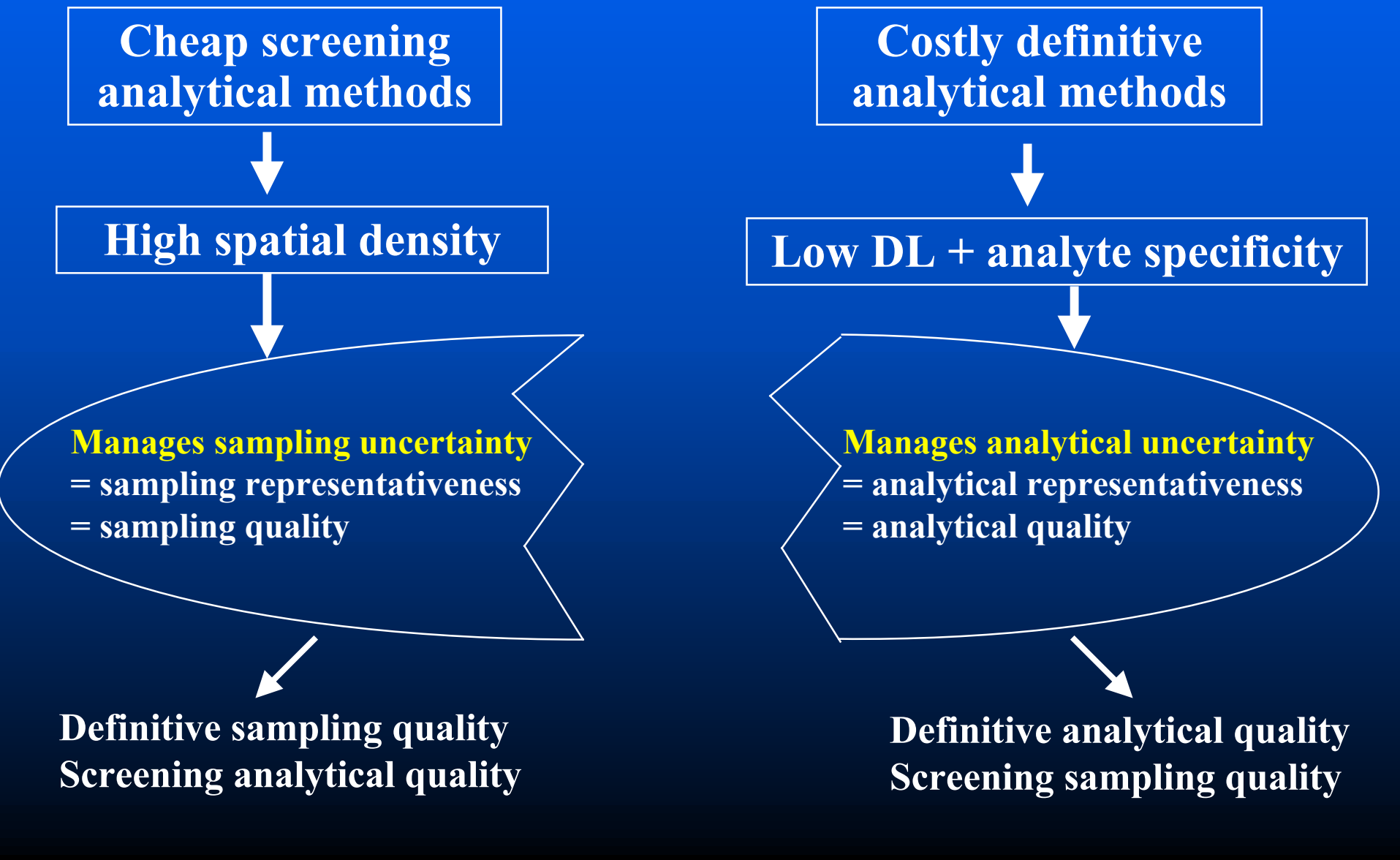
Decreased Sampling Variability after Removal of Hotspots

Ex 1

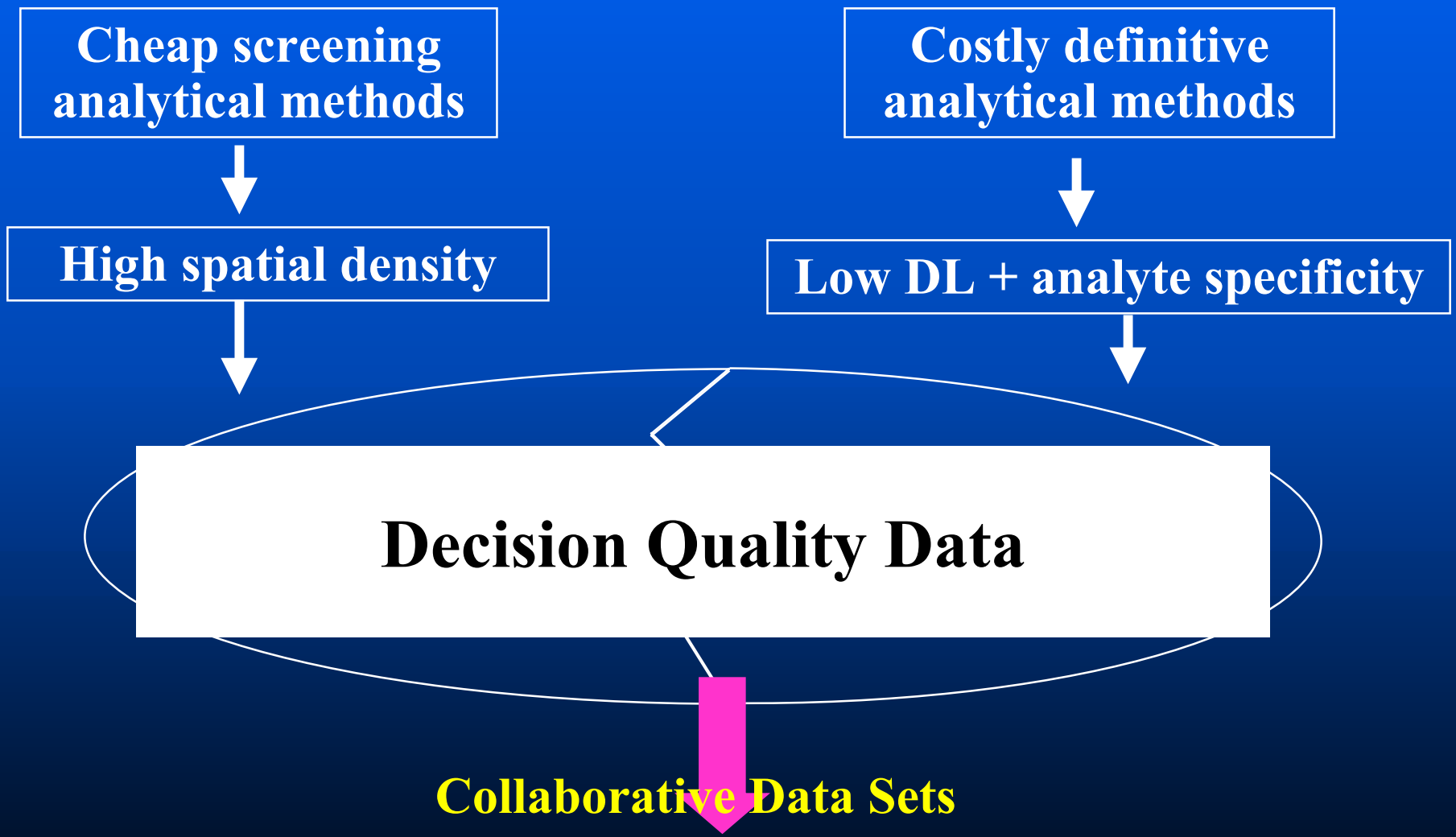
Ex 2

Ex 3

Marrying Analytical Methods to Make Sound Decisions Involving Heterogeneous Matrices



Marrying Analytical Methods to Make Sound Decisions Involving Heterogeneous Matrices



Case Study: Wenatchee Tree Fruit Site

- Pesticide IA kits guide dynamic work plan: 56 tons soil removed/segregate for incineration; 334 tons landfilled

230 IA analyses (w/ thorough QC) + 29 fixed-lab samples for 33 analytes

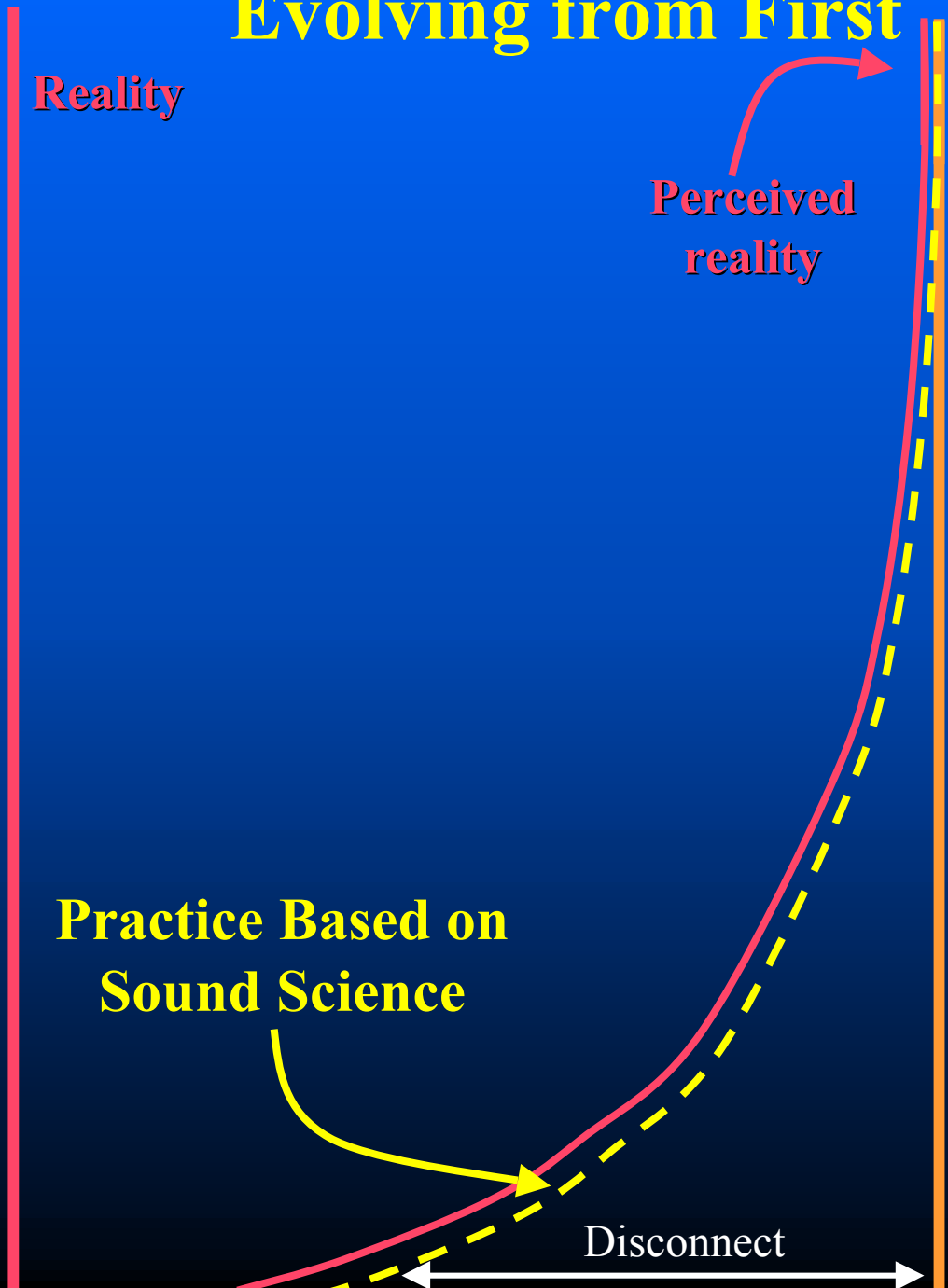
Managed sampling uncertainty: achieved very high confidence that all significant contamination was located and removed

Managed analytical uncertainty as additional QC on critical samples: confirm & perfect field kit action levels)

- Clean closure data set
  - 33 fixed lab samples for analyte-specific pesticide analysis
  - Demonstrate full compliance with all regulatory requirements for all 33 pesticide analytes to >95% statistical confidence
- Projected cost: ~$1.2M; Actual: $589K (Save ~ 50%)
- Field work completed: <4 months; single mobilization

So Where Do You Fit In?

Evolving from First Approximations



1982

Institutionalized Procedures, Program Guidance

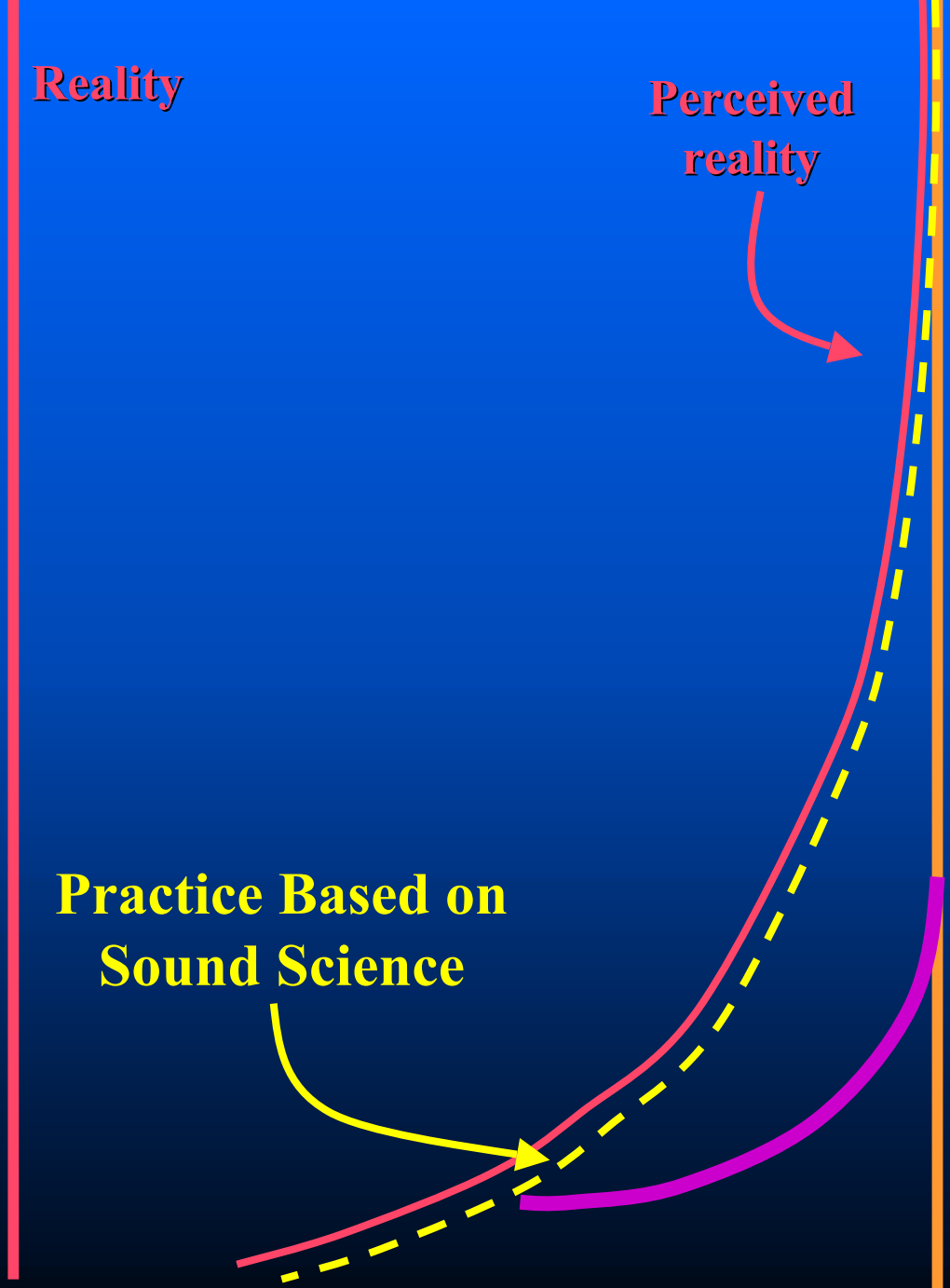
Time

Experience & investment in R&D produce

- Better technology tools
- More experience
- More complete knowledge
- Better models

Lagging Practices

Present

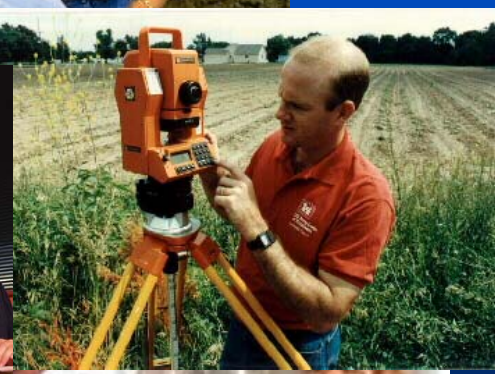

Reality
Perceived reality
Institutionalized Procedures, Program Guidance
Practice Based on Sound Science
Realign by controlled, gradual transition

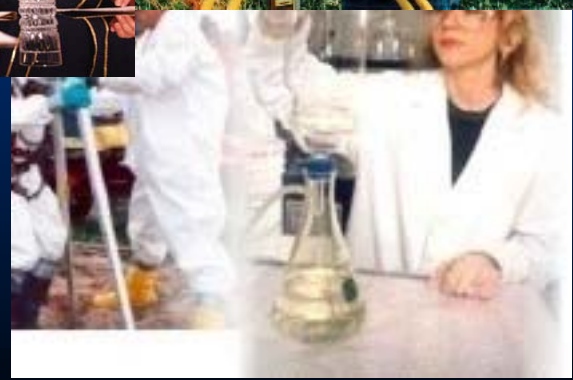
Adjustments to Facilitate "Sound Science"

- **Technical excellence through teaming**
- **Firmly ground site characterization and cleanup activities in project decision-making**
  - From: "define the nature and extent of contamination"
  - To: "define the nature and extent of contamination on the same scale as the scale of decision-making"
- **Implementation of sound science = Management of decision uncertainty (not solely dependent on data uncertainty)**
- **Revise EPA's data quality model (for contaminant chemistry data) to address all variables contributing to data-decision uncertainty**
  - Include sampling uncertainty in data quality
  - Report uncertainty interval in analytical quality
  - Revise detection/quantitation limit determination procedures

Chemist as Integral Part of Technical Team

- Assemble the project team by getting the right people involved
- May include statistician, chemist, hydro biologist, geol legal-regulatory adv etc.





"Data Quality" Terminology

Current terminology usage does not focus on the goal of decision quality

- Irony: Great focus on the quality of data points; but overall quality of decisions easily unknown.
- Current usage does not distinguish
  - Methods vs. data vs. decisions
  - The factors that impact each step in the process
  - Relationships between different aspects of quality

"Effective Data"
"Decision Quality Data"

Data of
known quality
that can logically be demonstrated to be
effective for making the specified decision
because both the
sampling and analytical uncertainties
are managed to the degree necessary to meet clearly
defined (and stated) decision confidence goals

Avoid Misleading Terminology

False Implications:

- **All methods run in the field are screening methods.**
- **All data produced in the field are of screening quality.**
- **Fixed labs using definitive analytical methods don't produce screening quality data.**
- **Fixed labs don't use screening methods.**

Proposed Clarification of Terms: QA

- **Project QA: ID causes of potential decision errors & the strategies to manage them**
  - DQOs are project decision goals; not data requirements!
- **Data QA: manage both sampling and analytical uncertainties to degree needed**
  - Analytical representativeness evaluated, including impact of sample/matrix effects on analytical performance
  - Sample representativeness evaluated
- **Lab QA: manage technical performance of analytical instruments, processes, and operators to meet lab proficiency goals**
  - Sample/matrix effects on analytical performance may or may not be evaluated—depends on contract specifications.

Proposed Clarification of Terms
Data Quality

- Decision quality data* = Effective data* = data shown to be effective for decision-making
- Screening quality data* = some useful information provided; but too uncertain to support decision-making alone
- Collaborative data sets = distinct data sets used in concert with each other to co-manage sampling and/or analytical uncertainties to an acceptable level

* Includes sampling uncertainty. Nature of method irrelevant.

Proposed Clarification of Terms Analytical QA Activities

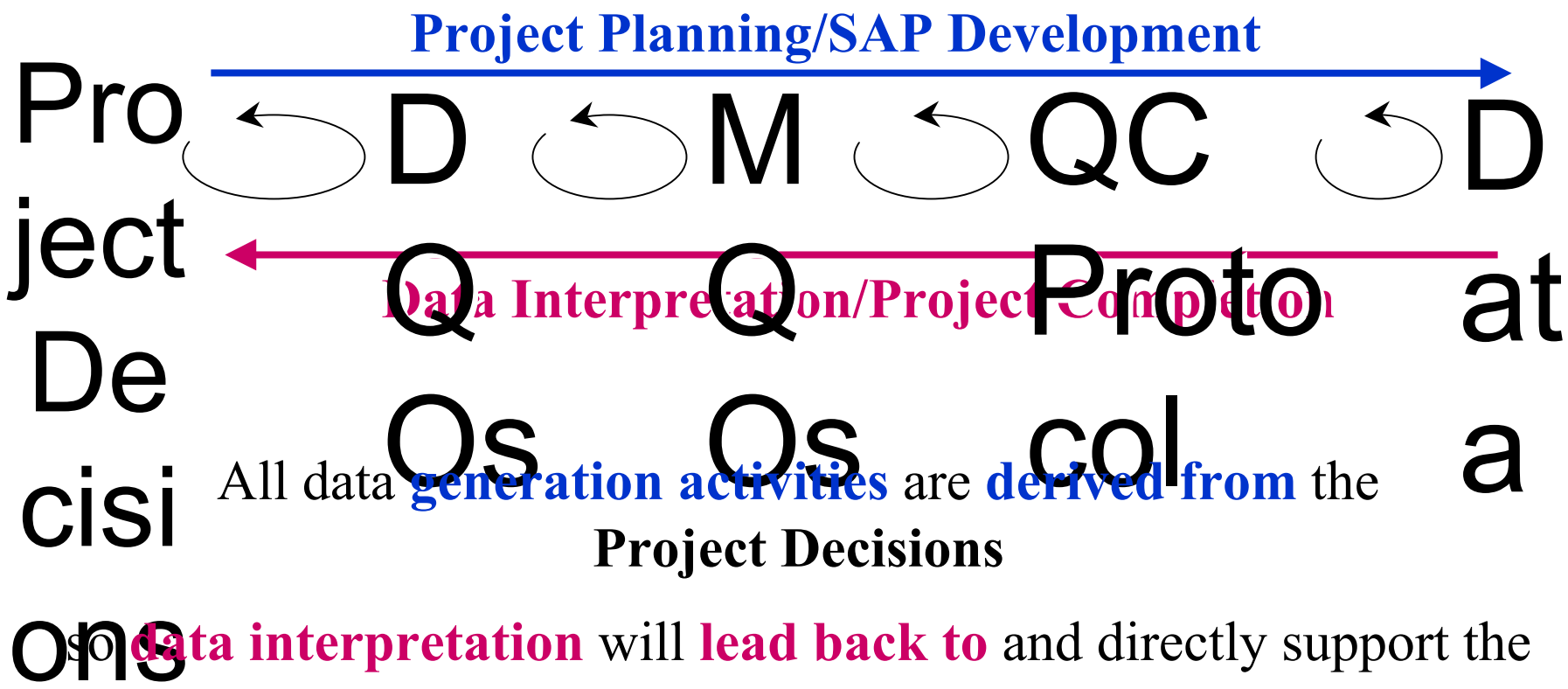
- Demonstration of method applicability
  - Shows that a particular method, project-specific SOP, and selected QC acceptance criteria are appropriate for a project-specific application or site-specific matrix
- Demonstration of proficiency
  - Shows that a particular operator or lab can perform a method properly

DQO Terminology

- DQOs: goal-oriented statements--establish technical bar for overall decision quality; express "what," not "how"
- MQOs: establish bar for data performance (may separate into sampling vs. analytical MQOs); "what," not "how"
- **Analytical QC acceptance criteria** (set after the "how" has been considered and selected):
  - Laboratory QC (monitor lab performance)
    - » Method-specific: lab equipment, lab procedures, analyst/operator
    - » May or may not monitor sample-specific impacts
  - Project QC (data of known quality to meet project goals)
    - » Method- & project-specific (requirements may be looser than lab)
    - » Must account for sample impacts

(See also the DQO Terminology paper)

DQO Term Relationships



All data **generation activities** are **derived from** the **Project Decisions**

so **data interpretation** will **lead back to** and directly support the **Project Decisions.**

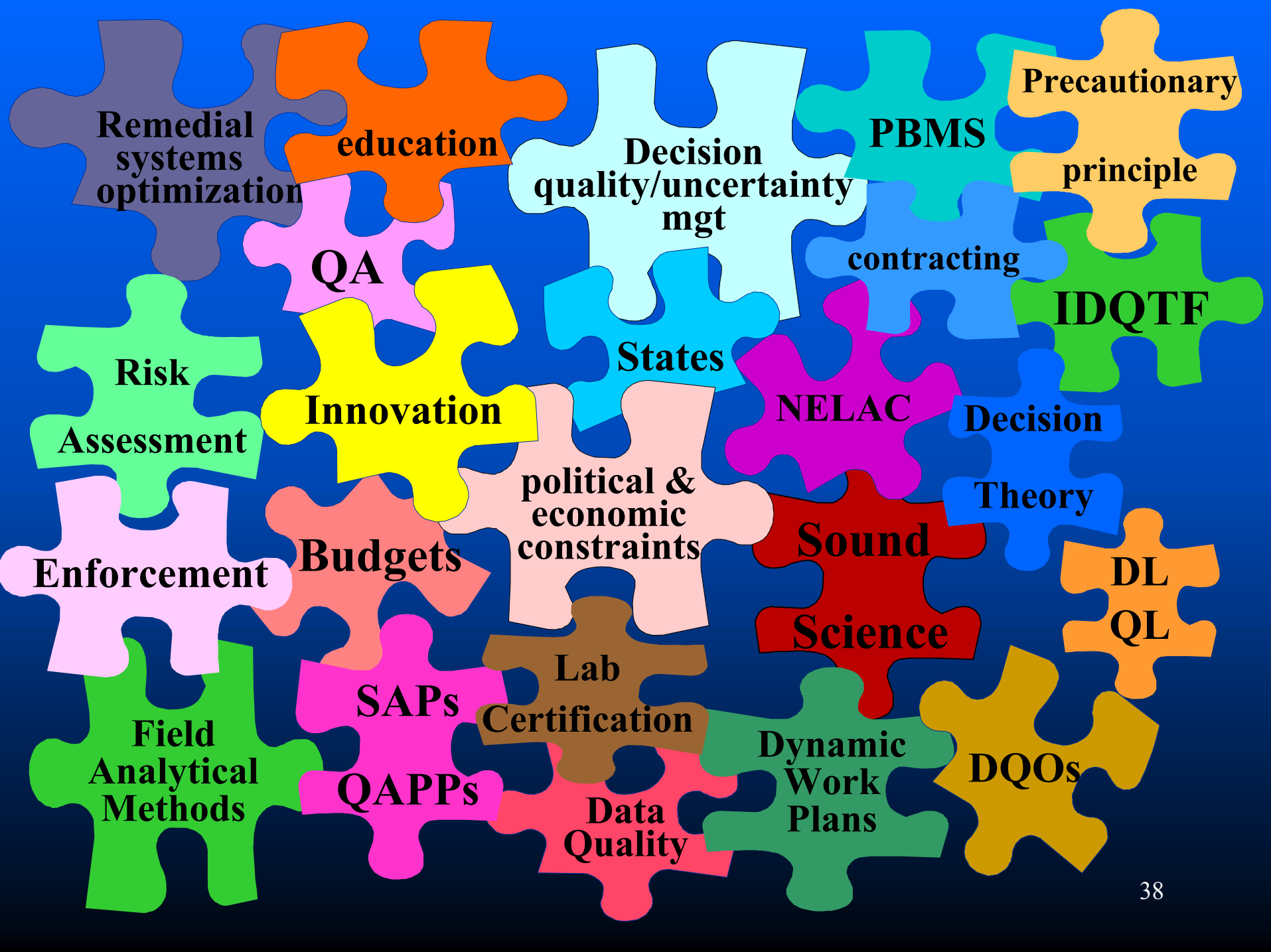
For more information, see the DQO Terminology paper (EPA 542-R-01-014) on webpage: http://cluin.org/tiopersp/issue.cfm

TQRSs

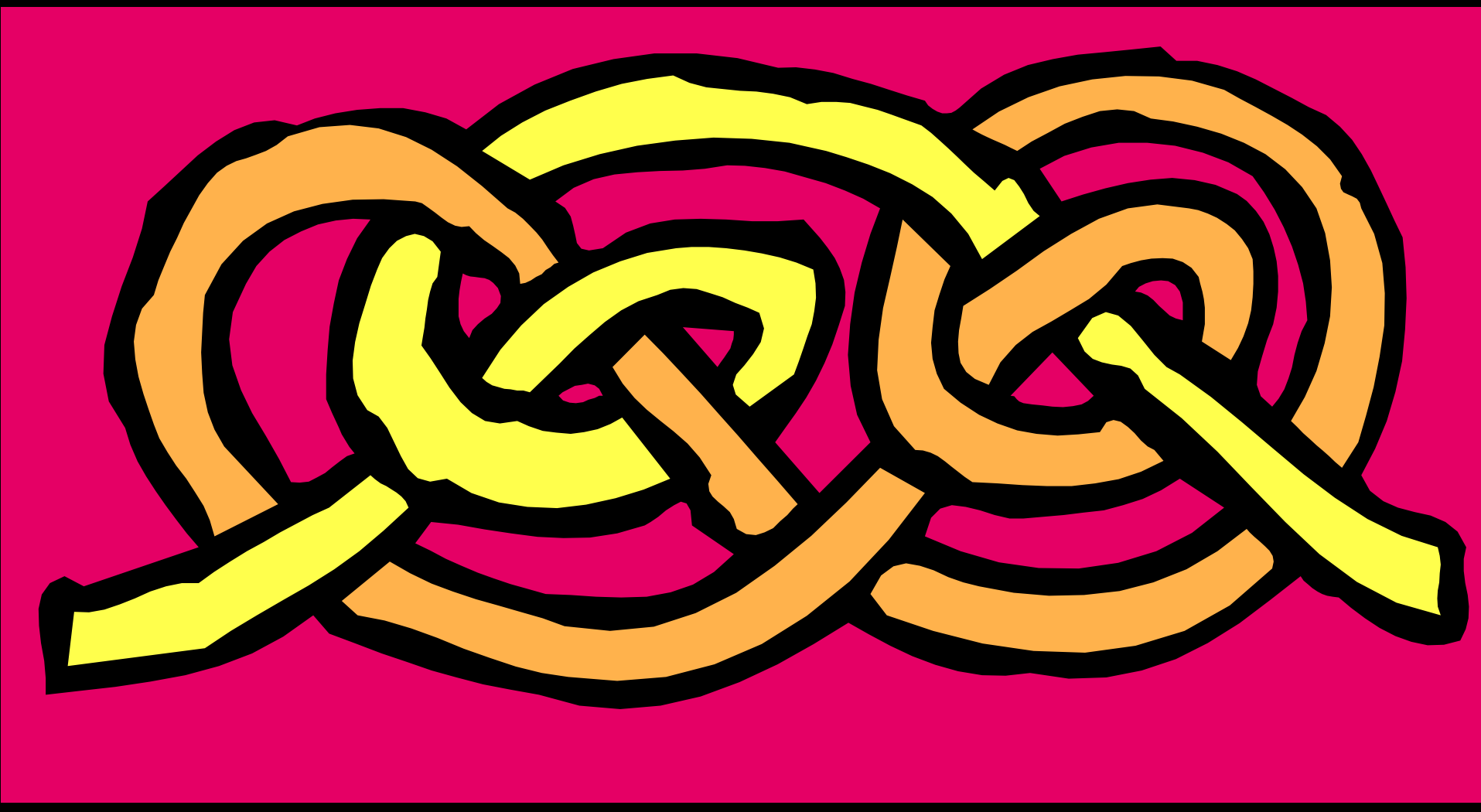
- Project Summary Forms and Technology Quick Reference Sheet (TQRS) template
- Use as a planning tool
  - Assist planning to generate data of known quality
- Use as a reporting tool
  - For rapid regulatory oversight of project basics
  - Documentation of analytical data quality
  - Compile 2-page cost & performance "case summaries" for projects using field analytical methods

TIO Efforts to Provide Support

- **Public outreach (also see Clu-In)**
  - Environmental Testing & Analysis article (Jan 2001)
  - ES&T feature article (Oct 2001; reprint available)
- **Project Managers Best Practices Handbook (in development—will need reviewers after July 31, 2002)**
  - Hyper-linked Internet-based "how-to" road map to existing EPA and other guidance and technical information w/ explanation to support implementation of Triad approach
- **Partnering with other organizations' efforts:**
  - US Army Corps of Engineers (Handbook partner)
  - Argonne National Lab (technical support)
- **Archived Internet seminars: http://cluin.org/studio/seminar.cfm**
- NEW! Chemists Corner: http://clu-in.org/chemistscorner/


Remedial systems optimization
education
Decision quality/uncertainty mgt
PBMS
Precautionary principle
QA
contracting
IDQTF
Risk Assessment
Innovation
States
NELAC
Decision Theory
political & economic constraints
Budgets
Sound Science
Enforcement
DL QL
Lab Certification
SAPs
Field Analytical Methods
QAPPs
Data Quality
Dynamic Work Plans
DQOs

Putting all the Pieces Together: Manage Decision Uncertainty



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