User Guide

Uniform Federal Policy
Quality Assurance Project Plan Template
For
Soils Assessment of Dioxin Sites

September 2011

ACRONYMS

ARAR	Applicable or Relevant and Appropriate Requirement
CALUX	Chemical Activated Luciferase Gene eXpression
CERCLA	Comprehensive Environmental Restoration, Compensation, and Liability Act
CLP	Contract Laboratory Program
COC	Contaminant of Concern
CRREL	Cold Regions Research and Engineering Laboratory
CSM	Conceptual Site Model
CV	Coefficient of Variation
DL	Detection Limit
DMA	Demonstration of Method Applicability
DNT	Dinitrotoluene
DO	Dissolved Oxygen
DQA	Data Quality Assessment
DQI	Data Quality Indicator
DQO	Data Quality Objective
DU	Decision Unit
DWS	Dynamic Work Strategies
EA	Exposure Area
EMPC	Estimated Maximum Possible Concentration
EPA	U.S. Environmental Protection Agency
EPC	Exposure Point Concentration
ESTCP	Environmental Security Technology Certification Program
EU	Exposure Unit
FP-XRF	Field Portable X-Ray Fluorescence
GIS	Global Imaging System
GPS	Global Positioning System
HHRA	Human Health Risk Assessment
HQ	Headquarters
ICS	Incremental Composite Sampling
ISM	Incremental Sampling Methodology
ITRC	Interstate Technology and Regulatory Council
KM	Kaplan-Meier
mm	Millimeter
MS	Matrix Spike
MSD	Matrix Spike Duplicate
MS/MSD	Matrix Spike/Matrix Spike Duplicate
MSD	Matrix Spike Duplicate
NFA	No Further Action
ORD	Office of Research and Development
OSRTI	Office of Superfund Remediation and Technology Innovation
PAH	Polycyclic aromatic hydrocarbons
PCB	Polychlorinated biphenyl
PCDD	Polychlorinated dibenzo-p-dioxin

PCDF	Polychlorinated dibenzofuran
ppm	parts per million
ppt	parts per trillion
PQO	Project Quality Objective
PRG	Preliminary Remediation Goal
QA	Quality Assurance
QA/QC	Quality Assurance/Quality Control
QAPP	Quality Assurance Project Plan
QC	Quality Control
RAGS	Risk Assessment Guidance for Superfund
RSD	Relative Standard Deviation
RPM	Remedial Project Manager
RSD	Relative Standard Deviation
SD	Standard Deviation
SOP	Standard Operating Procedure
SOW	Statement of Work
SSG	Soil Screening Guidance
SSL	Soil Screening Level
SU	Sampling Unit
TCDD	Tetrachlorodibenzo-p-dioxin
TEF	Toxicity Equivalency Factor
TEQ	Toxicity Equivalent
TIIB	Technology Integration and Information Branch
TIFSD	Technology Innovation and Field Services Division
TNT	Trinitrotoluene
TOC	Total Organic Carbon
UCL	Upper Confidence Limit
UFP-QAPP	Uniform Federal Policy-Quality Assurance Project Plan
USACE	United States Army Corps of Engineers
USEPA	United States Environmental Protection Agency
VOC	Volatile Organic Compound
VSP	Visual Sampling Plan
WHO	World Health Organization
XRF	X-Ray Fluorescence

User Guide
Uniform Federal Policy - Quality Assurance Project Plan Template for Soils Assessment of Dioxin Sites

Table of Contents

Title **Page**

ACRONYMS i
1.0 INTRODUCTION 1
2.0 PREPARATION OF SELECT UFP-QAPP TEMPLATE SECTIONS 5

EXHIBITS

Exhibit 1: Figure 1: Decision Logic Diagram for Dioxin Site Assessment Process 4
Exhibit 2: Attachment A to Figure 1 - Historical Documentation and Data Review 6
Exhibit 3: Attachment B to Figure 1 - Systematic Planning Meeting Agenda[s] 10
Exhibit 4: Attachment C to Figure 1 - CSM and Data Gap Assessments During Systematic Planning 14
Exhibit 5: Attachment D to Figure 1 - Background Study Design and Performance 18
Exhibit 6: Attachment E to Figure 1 - Summary of Assessment Effort, Objectives and Assumptions 23
Exhibit 7: Attachment F to Figure 1 - TEQ-Based Risk Screening 24
Exhibit 8: Attachment G to Figure 1 - Conducting A Human Health Risk Assessment 25

FIGURES

Figure 1. Single DU with 30 Increments Going into a Single ICS Sample 30
Figure 2. Three Replicate DU-ICS Samples of 30 Increments Each 30
Figure 3. Example of Four SUs Within a DU 33
Figure 4. Variability Source QC Procedure: Measures Sources of Data Variability 38
Figure 5. Example of Random Distribution of Contamination 39
Figure 6. Example of Spatially Stratified Distribution of Contamination 40
Figure 7. Example of Contaminant Distribution Overtime 67

APPENDICES

1: Best Practices Inventory for Implementing and Assessing Elements of a Successful Project
2: Rationale for a Default of 30 – 60 Increments
3: Demonstrations of Methods Applicability under a Triad Approach
4: Calculation of Total Dioxin TEQs with Nondetect and Rejected Congeners
5: References

User Guide
Uniform Federal Policy - Quality Assurance Project Plan Template for Soils Assessment of Dioxin Sites

1.0 INTRODUCTION

This User Guide was prepared by the U.S Environmental Protection Agency (EPA) Office of Superfund Remediation and Technology Innovation (OSRTI) to provide explanations and instructions for use of the Uniform Federal Policy – Quality Assurance Project Plan (UFP-QAPP) template for shallow soils assessment of dioxin sites. The purpose of the User Guide is to support development of project-specific QAPPs based on the UFP-QAPP template. The User Guide provides context, examples, and detailed explanation; however, it is not intended to address every section of the UFP-QAPP. The act of completing a project-specific UFP-QAPP, however, is intended to guide the project team through the systematic planning effort needed for a successful dioxin assessment effort.

Tasks or issues that are not unique to dioxin assessment data collection projects, such as community involvement and access agreements, are not treated in-depth in the UFP-QAPP template or User Guide.

An extensive companion UFP-QAPP Manual (March 2005) describing the design and use of the UFP-QAPP, as well as all other original UFP-QAPP associated documents and templates, can be downloaded for reference from http://www.epa.gov/fedfac/documents/qualityassurance.htm. The original UFP-QAPP Manual should be considered a valuable companion to this User Guide. For some topics, much more detail is provided in the UFP-QAPP Manual than is presented herein.

Regions have the discretion to assess dioxin sites in the manner they determine best addresses site-specific situations and concerns. The goal of the UFP-QAPP template and User Guide however, is to provide a consistent approach to dioxin assessments with a primary focus on protecting human health and the environment and doing so in the context of managing site uncertainties and resources to accomplish assessment efforts. The strategies presented in this User Guide are based on incremental and compositing techniques for soils. Both compositing and incremental sampling refer to the same basic process of mixing portions together, however, the purpose and details differ. EPA's "compositing" term is broader in its goals and applications than incremental sampling, for example, there are some composite designs used for hot spot searching. Incremental sampling, a more recent term, however, is considered a specific type of compositing used to derive an average. For the purpose of this User Guide and the accompanying UFP-QAPP template, it is recognized that sampling designs may include a number of objectives. Since one or more of these goals might be present in a single sampling design, the term "incremental-composite sampling" (ICS) is used herein to cover all possible sampling goals.

This User Guide presents the designs and rationale behind several ICS assessment strategies for soils, in order of increased level of complexity, as follows:

- DU-Based Default Strategy = sampling of whole decision units (DUs) using ICS, with the ICS samples comprised of a uniform range of 30 to 60 sample increments.

- SU/DU-Based Modified Strategy = dividing DUs into equally-sized sampling units (SUs) and sampling each SU using ICS samples comprised of equal numbers of sample increments that in total for the DU are within the range of 30 to 60. The optional purposes for using this strategy include:
    - Facilitating efficient collection and archiving of sub-DU (i.e, SU) samples for potential analysis to provide sub-DU information in support of remedial planning;
    - According to the specifics of the conceptual site model (CSM), accommodating heterogeneity of contaminant distribution attributable to sub-DU scale physiographic site features; and
    - Identifying and managing other drivers of short- and long-scale heterogeneity.

- Statistically-Based Strategy = using statistical and CSM information from a project-specific pilot study to determine how to appropriately divide DUs into SUs, and calculate how many SUs to sample and how many increments to use to comprise ICS samples. The optimal purposes for using this strategy include:
    - Dealing with very large areas needing assessment;
    - Dealing with very large DUs; and/or
    - Justifying the use of less than 30, or more than 60, increments per DU.

The goal of the UFP-QAPP template is to ensure quality performance of a dioxin site assessment project. Completing a project-specific UFP-QAPP for each site assessment serves to ensure that the project and its documentation meet the requirements of USEPA Quality policies (see page vii of the original UFP-QAPP Manual (March 2005)). The UFP-QAPP template was designed to be adaptable and flexible to serve site-specific circumstances. The UFP-QAPP template worksheets are intended to be filled out during systematic planning meetings. Additional discussions or efforts needed to complete a project-specific UFP-QAPP should be performed prior to completion of systematic planning for the project.

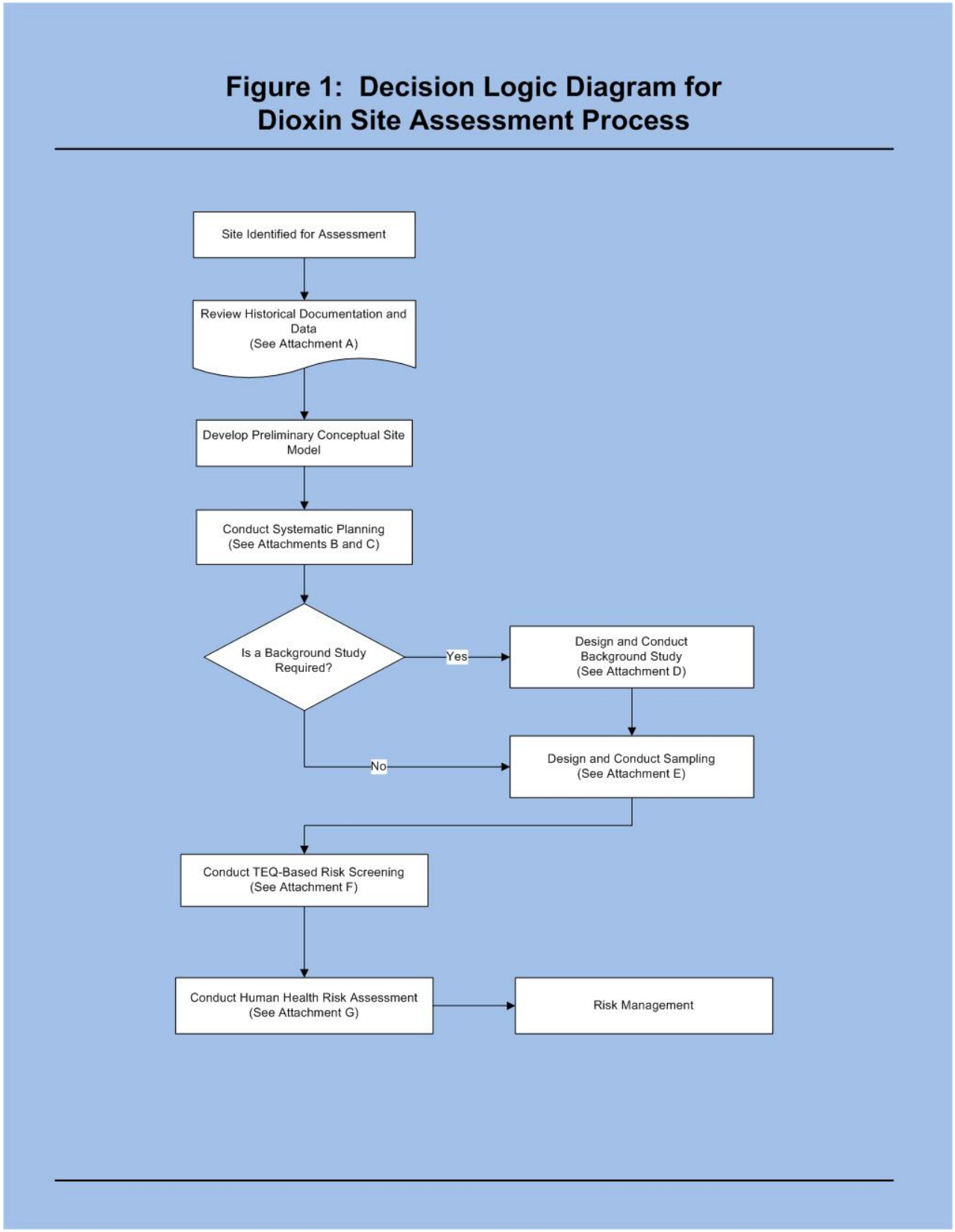
Writers of project-specific UFP-QAPPs should become familiar with the worksheets, the original UFP-QAPP Manual, and this User Guide prior to initiating a dioxin site soils assessment project. Worksheets may appear to follow a different order from the sequence of activities through time. Personnel involved in project-specific UFP-QAPP preparation should expect an iterative process of planning activities and recording outcomes to ensure consistency with other UFP-QAPP components. It may be necessary to consider a given worksheet multiple times in an effort to refine its content as the project goals and work strategies take shape over the course of the planning effort.

Based on project needs, writers may also need to prepare multiple copies of some worksheets. For instance, data collection worksheets may be necessary for both a background study and assessment efforts, or where there are variations in approaches required for one or more of a multiple DU-based project. Regions may discover that some activities may require development of entirely new worksheets or attachments to add to the site-specific UFP-QAPP. Any worksheets determined to be irrelevant to a given project should be marked with an explanation for non-use to prevent field staff and other readers from unnecessarily questioning document completeness.

Applicable generic text and information prompts are provided throughout the UFP-QAPP template to expedite site-specific UFP-QAPP development. Some prompts require specific information, whereas others serve only as examples of the type of information requested. The examples provided are not mandatory, nor intended to be a comprehensive or complete list. Planning teams are responsible for determining appropriate content, thus an expectation is that some prompts may be added, deleted or modified as a function of Region-, site- and project-specific planning efforts. Regardless of the Regional decision-making process used for a particular site, the primary goal of a project-specific UFP-QAPP is to capture and transparently document the actual process to be used.

Exhibit 1 shows a flow chart of the typical process anticipated for a dioxin-assessment project.

Exhibit 1: Figure 1: Decision Logic Diagram for Dioxin Site Assessment Process



2.0 PREPARATION OF SELECT UFP-QAPP TEMPLATE SECTIONS

This section provides explanations for preparing select UFP-QAPP template sections.

Review Historical Information and Data – This entry is intended to be used to document the beginning of the quality assurance (QA) process by affirming that the project team has reviewed relevant site historical information and data. This review of historical material serves as the basis for developing a Preliminary Conceptual Site Model (CSM) for the site and property (ies) that are to be assessed. The Preliminary CSM is a critical element of the systematic planning process as it summarizes all the information readily available at the beginning of a project prior to the systematic planning effort (USEPA, 1996a; see section 2.1 for more details).

Attachment A to Figure 1 – This entry prompts a reviewer for items to consider when reviewing historical data. The Preliminary CSM is the recommended way of documenting the reviews of historical information and data, along with the conclusions drawn from them. The summary results of the data quality assessment of historical data should be recorded on UFP-QAPP Worksheet #13 (the usability of secondary data). Attachments to Worksheet #13 should record the detailed findings prompted in Attachment A to Figure 1. For convenience, a copy of the contents of Attachment A to Figure 1 follows on the next page.

In the UFP-QAPP template, documentation of historical information and data review begins in the Introduction section: "Between xx/xx/201x and yy/yy/201y [insert DATES]" – refers to the development of the Preliminary CSM. It is important to note that the need for some historical documentation may be identified during a systematic planning effort; however, the Preliminary CSM should be as complete as possible prior to the formal systematic planning effort to capture available information and identify potential data gaps. Exactly how this part of the project progresses is at the discretion of the Remedial Project Manager (RPM).

Exhibit 2: Attachment A to Figure 1 - Historical Documentation and Data Review

Attachment A to Figure 1
Site Assessment Sampling Design and Strategy

Uniform Federal Policy (UFP) Quality Assurance Project Plan (QAPP)

This section provides examples of considerations for reviewing site historical information in light of a proposed dioxin assessment. Information documented in a project-specific UFP-QAPP using the accompanying template should describe what documents or information was actually reviewed and what conclusions or work products (maps, figures, etc.) resulted from these reviews.

Historical Documentation and Data Review

Historical site documentation and data should be compiled and reviewed by the project team. The results of this effort serve as the basis for developing the Preliminary CSM and informing the systematic planning effort. Historical site data sets should be evaluated for applicability as data needed to aid assessment against the soil Preliminary Remediation Goal (PRG) for dioxin toxicity equivalents (TEQ)s. While being compiled, the quality and usability of historical data should be assessed from both sampling and analytical perspectives. Assessment of existing information and conclusions drawn based on historical sampling and analytical methods should be documented in the UFP-QAPP. Documentation can be summarized in Worksheet #13 (for the usability of secondary data), with attachments to Worksheet #13 added to record any detailed findings. Data quality assessment addresses the following items (more detail is available in EPA's data usability guidance (USEPA, 1991)).

Evaluation of Historical Sampling Approach

- General sampling strategy
  - Was there a basis for the sampling design in the CSM of that time?
  - Was there a rationale for the choice of sampling design?
  - Was it a statistical/probabilistic or judgmental design? Describe the design.
  - Was a background study conducted?

- Sample representativeness and comparability relative to new data needs
  - Soil media sampled (sites and sub-sites, soil/waste types, background vs. site)
  - Sampling density
  - Depth intervals
  - Grab or ICS samples
  - Sample processing (sizing, homogenization)

- What was the intended use of the historical data?
  - Site delineation or screening
  - Risk assessment
  - Remedial design/remedial action (engineering evaluations, characterization of treated or removed wastes, confirmation of soil/waste removal)

- Decision uncertainty management approach
  - Qualitative/professional judgment
  - Analytical quality assurance (QA) program only
  - Classical statistics
  - Other (e.g., geostatistics, modeling)
  - Unknown

Data Quality Assessment via Evaluation of Analytical Methods and Quality Assurance Program

- Known and documented quality of data; i.e., were samples analyzed and reported as well as validated under an EPA QA program or equivalent?
- Status of analytical data in terms of whether it was collected for all contaminants of concern (COCs) for use in the current TEQ evaluations (dioxins, furans, and dioxin-like polychlorinated biphenyls (PCBs))
- Were quantitation/detection limits sufficient for use in the current Toxicity Equivalents (TEQ) evaluations?
- Did data quality indicators (DQIs; i.e., quality control (QC) checks for precision, matrix interference, etc.) meet method performance requirements and did they indicate sufficient data quality for use in the current TEQ evaluations (e.g., precision, bias, completeness, comparability)?
- Were there any applications of field-based or screening methods (e.g., CALUX or immunoassay methods)?
- If non-conventional methods were used, was there a demonstration of method applicability (DMA) or other type of pilot study, or subsequent data analysis to establish the comparability between conventional and alternative methods?
- Did any of the historical analytical methods find matrix interferences that warrant consideration when selecting extract cleanup methods for future analyses?
- Are there QC or validation records available for any applications of non-conventional methods?

Ideally, a thorough analysis of historical data would determine if previous data could be used to guide assessment planning, or in some cases even provide data of adequate quantity and acceptable quality to offset some of the assessment requirements. If of adequate quality, these data might be used to:

- suggest optimal sampling and analytical strategies,
- help the project technical team to develop appropriate size, shape, and orientation of assessment decision units (DUs)
- support determination of constituent background concentrations,
- substitute for, or augment, current data collection needs,
- perform TEQ-based risk screening by evaluating the total TEQ value for all dioxin-like compounds against the PRG,
- perform human health risk assessment (HHRA), and
- provide information for remediation / mitigation planning and engineering.

If the sampling and analytical quality of historical data is undocumented and unknown, its use is greatly restricted. Possibly this data can be helpful in developing a preliminary CSM and potentially applicable remedial strategies, however this CSM should be carefully tested for accuracy during project implementation. If of known quality, however, historical data may still have much value even if the data are not adequate to substitute for current data needs. For example, although quantitation limits might not be adequate for current needs, the data may still provide valuable qualitative information to support confident development of the Preliminary CSM. The historical data may also indicate how much variability exists in TEQ concentrations in the field. An understanding of this variability can help provide the basis for determining the number of increments appropriate to address this level of heterogeneity. For example, variability associated with historical data may indicate it is appropriate to increase the number of increments within the default value range (30 to 60 increments) from a number around 30 to a number closer to 60 in an effort to ensure accurate representation of the mean.

Systematic Planning – A short summary documenting the multi-disciplinary, team-based systematic planning effort conducted should be provided in Attachment A to Figure 1 of the project-specific UFP-QAPP. The project-specific UFP-QAPP is the primary written product of the systematic planning effort. It serves both to document the planning process and as a guide to prompt the project team through systematic planning for the assessment effort. A detailed "Best Practices" checklist is included in Appendix 1 of this User Guide. Use of this checklist is not a requirement, but it has several potential beneficial uses, such as to assist:

- general project planning by helping to ensure that key elements are not overlooked during development of project plans,
- reviewers providing project QA by ensuring that all the topics supposed to be covered in the project-specific UFP-QAPP are present, and
- project assessment and documentation; by providing a record of what activities were actually conducted and goals accomplished over the course of the project.

The outcome of project-level assessments (as opposed to data assessments) should be recorded on Worksheet #32 in the project-specific UFP-QAPP. Eight varieties of project-level assessments, including "readiness reviews" and "technical systems audits" are provided in the original UFP-QAPP Manual (March 2005); see page 100. Project-level assessments should compare project implementation to what was prescribed in the project-specific UFP-QAPP to ensure that the project team's goals, as they were determined through the systematic planning process, were met. Any significant deviations from the project-specific UFP-QAPP's instructions should be recorded in field logbooks, along with the reason and the substituted procedures. The field crew's logbook notations are attached to Worksheet #32 assessments. The project team may also choose to instruct the field team to notify them right away and seek approval before implementing a deviation in procedures. If such a deviation is reasonably anticipated, project teams can include this instruction in the relevant standard operating procedure (SOP) and worksheet, as well as on Worksheet #6 ("Communication Pathways"). Anticipating contingencies and having procedures in place to achieve consensus on potentially important deviations are critical parts of a well-planned project.

A primary goal of the UFP-QAPP template and systematic planning is to provide a cost-effective sampling strategy that meets data objectives in a limited number of field mobilizations. A comprehensive process can help limit duplication of efforts and the need for multiple mobilizations arising from insufficient planning. Differing views among stakeholders that could potentially cost time and resources are best to address prior to initiating field efforts. Experience has proven that systematic planning is the most effective mechanism for developing a successful project. The project team should resist pressures to bypass or employ inadequate time and resources for the systematic planning process. USEPA HQ staff is available to provide information and technical assistance to the RPM, risk assessor, or other team members. Requests for technical assistance may be submitted via the OSRTI Technology Integration and Information Branch (TIIB) by contacting Dan Powell, the Branch Chief for TIIB, via email at powell.dan@epa.gov or via phone at 703-603-7196. Note that all participating project staff (including contractors) and their qualifications need documentation in the project-specific UFP-QAPP on Worksheets #7 and 8. Project-specific UFP-QAPP Worksheets #1, 3, 5 and 6 should be used to document other organizational QA activities and the roles and responsibilities of project staff and higher levels of organization management.

Systematic planning involves identifying project decisions in clear and unambiguous language, along with addressing any special considerations and complications for a particular site. It also engages stakeholders in gathering historical information and identifying concerns that drive the framing of future project decisions. Stakeholders should be included through all key decision-making stages.

Planning includes building in contingencies to accommodate changes in project conditions. This requires staff to anticipate what could go wrong during project implementation. If feasible, the project team should plan proactive measures to prevent problems that have a high probability of occurring or significant potential impact to a project. These discussions allow the project team to have some contingencies in place for issues with increased likelihood of occurrence and significant impact to project goals or schedule while outlining a process for resolving unanticipated issues in a timely fashion. Contingency discussions, at the very least, enable the project team to be prepared in the event that a significant project issue arises.

An example systematic planning meeting agenda is provided as Attachment B to Figure 1. UFP-QAPP template Worksheet #9 should be used to document planning meeting participants. Meeting minutes, with action items and significant agreements reached, should be included in the project-specific UFP-QAPP as attachments to Worksheet #9.

A variety of EPA guidance can be consulted to provide additional detail to support the systematic planning process, including

- EPA Quality Manual for Environmental Programs. (USEPA 2000, May),
- Guidance on Systematic Planning Using the Data Quality Objectives Process. (USEPA 2006, February), and
- Guidance for Developing Quality Assurance Project Plans. (USEPA 2002a, December)

Exhibit 3: Attachment B to Figure 1 - Systematic Planning Meeting Agenda[s]

Attachment B to Figure 1
Site Assessment Sampling Design and Strategy
DRAFT Uniform Federal Policy (UFP) Quality Assurance Project Plan (QAPP)

This section provides examples of agenda items and considerations for systematic planning efforts. Information documented in a project-specific UFP-QAPP using the accompanying template should describe team members who participate, project goals/exit strategies, key site decisions and project metrics, as well as the agreed upon technical approach to components of the project-specific dioxin assessment. Systematic planning meetings provide the technical team with an opportunity to agree upon and document the dioxin assessment process by filling out UFP-QAPP template worksheets and attachments.

Systematic Planning Meeting – An Example Agenda

- Team member introductions
  - Meeting roles and expectations
  - Lines of project authority and communication

- Confirm general project goals, site problems, and exit strategy

- Confirm site reuse goals (residential, industrial, recreation, other)

- Identify key site decisions:
  - Confirm site contaminants of concern (COCs)
  - Evaluate risk assessment inputs such as potential pathways and receptors
  - Confirm soil screening criteria
    - Preliminary Remediation Goals (PRGs)
    - Background
    - Applicable or Relevant and Appropriate Requirements (ARARs)
  - Definition of completed delineation

- Develop key project metrics (schedule, budget, other)

- Update Preliminary CSM to Baseline CSM status with the team if needed. The Baseline CSM represents the project technical team's understanding and agreement of existing data needs, potential data gaps, site geologic/hydrogeologic features, etc. formulated from historical data review and the preliminary CSM.
  - Identify key data gaps
  - Identify missing elements, new hypotheses, etc

- Discuss technical approach to assessment
  - Project-specific UFP-QAPP - Fill out the worksheets thoroughly and completely.
  - Develop decision criteria, decision-making processes, decision logics, rules, etc.

- Data collection - start with a general approach that gets more specific as planning unfolds over the course of one or several meetings
  - Sampling design
  - Decision units (DUs)
- Analytical methods - consider the benefits and limitations of both conventional and innovative methods (e.g., CALUX and immunoassays). Consider that innovative methods may possibly provide important CSM and heterogeneity information with rapid turnaround of results, even if not having all the sensitivity capabilities of conventional methods.
- Develop approaches to measure and manage both sampling and analytical uncertainties.
- Identify technologies and methods that may require demonstrations of method applicability (DMA) to establish method comparability, usefulness, and gain expertise in their use.
- Where useful and feasible, plan for real-time data management, assessment, visualization and communication.

- Meeting Summary and documentation of key decisions and strategies.

Post-Meeting Activities

- Prepare meeting minutes and circulate to stakeholders.
- Write up documentation of agreements and areas of consensus. Secure signatures documenting the agreement of participants/stakeholders; maintain those documents in the project-specific UFP-QAPP records.
- Follow-up discussions on incomplete agreements as needed – the ideal goal is for all stakeholders to be in agreement/consensus on issues, or in agreement on a process to resolve issues identified during systematic planning.
- Technical team performs additional research as needed on potential characterization and/or remediation technologies, designs a DMA/pilot study if appropriate, and explores various sampling designs. Results should be presented for discussion and consensus, including at a subsequent systematic planning meeting, as warranted.
- Finalize additional data gap identification for the CSM as needed.
- Update CSM to Baseline status in designated documentation and visualization formats.
- Review the project-specific UFP-QAPP material added during the meeting; ensure there is clarity and consistency in the writing; identify items for follow-up and additional clarification, including at a subsequent systematic planning meeting, as warranted.
- Plan for procurement of technologies and services.

UFP-QAPP template Worksheet #9 should be used to document planning meeting participants. Meeting minutes, with action items and significant agreements reached, should be included in the project-specific UFP-QAPP as attachments to Worksheet #9.

Developing the Baseline CSM from the Preliminary CSM

The Preliminary CSM is a representation of contamination concerns that are known, or suspected, to be present; as well as an aid to predicting the nature, exposure, and extent of the contamination. It can be expressed through text, tabular data and/or simplified graphic renderings; or more complex visualization tools in order to capture, communicate, and leverage existing information. The Preliminary CSM enhances stakeholder understanding of site conditions and helps to focus future investigation and remediation efforts on key uncertainties or data gaps. CSM development, project goals, and supporting information should be documented as attachments to Worksheet #10.

The Preliminary CSM should be distributed to stakeholders for review prior to the first systematic planning meeting. [See USEPA, 1993 and ITRC, 2003 for more information about constructing CSMs.] The Preliminary CSM provides systematic planning participants with a basis for discussing current project objectives in the context of available historical information. As the project team works during systematic planning meetings, the Preliminary CSM is evolved to a Baseline CSM as historical data is explored; updates to site graphics and maps are developed; and deposition, transport/migration and exposure mechanisms are proposed, and site reconnaissance confirms current site conditions. It is important to seek stakeholder input and concurrence as the Baseline CSM evolves (USEPA, 1996a, 2002c). The Baseline CSM is then used to support detailed planning of the sampling program and the logistics and sequencing of the field project, including:

- identifying information gaps and the data needed to fill them,
- developing risk assessment priorities and pathway-receptor network diagrams,
- developing the site-specific sampling plan design,
- refinement of the strategy for data collection, statistical analysis, and data use in the decision-making process,
- addressing areas of stakeholder contention, and
- anticipating sources of data variability that could interfere with decision-making, such as the degree of matrix and spatial heterogeneity, sub-sampling variability, and the analytical method variance.

Note: The CSM concept used in these documents, and in accordance with USEPA guidance, is not limited to creation of a pathway-receptor network diagram.

Planning Based on the Baseline CSM

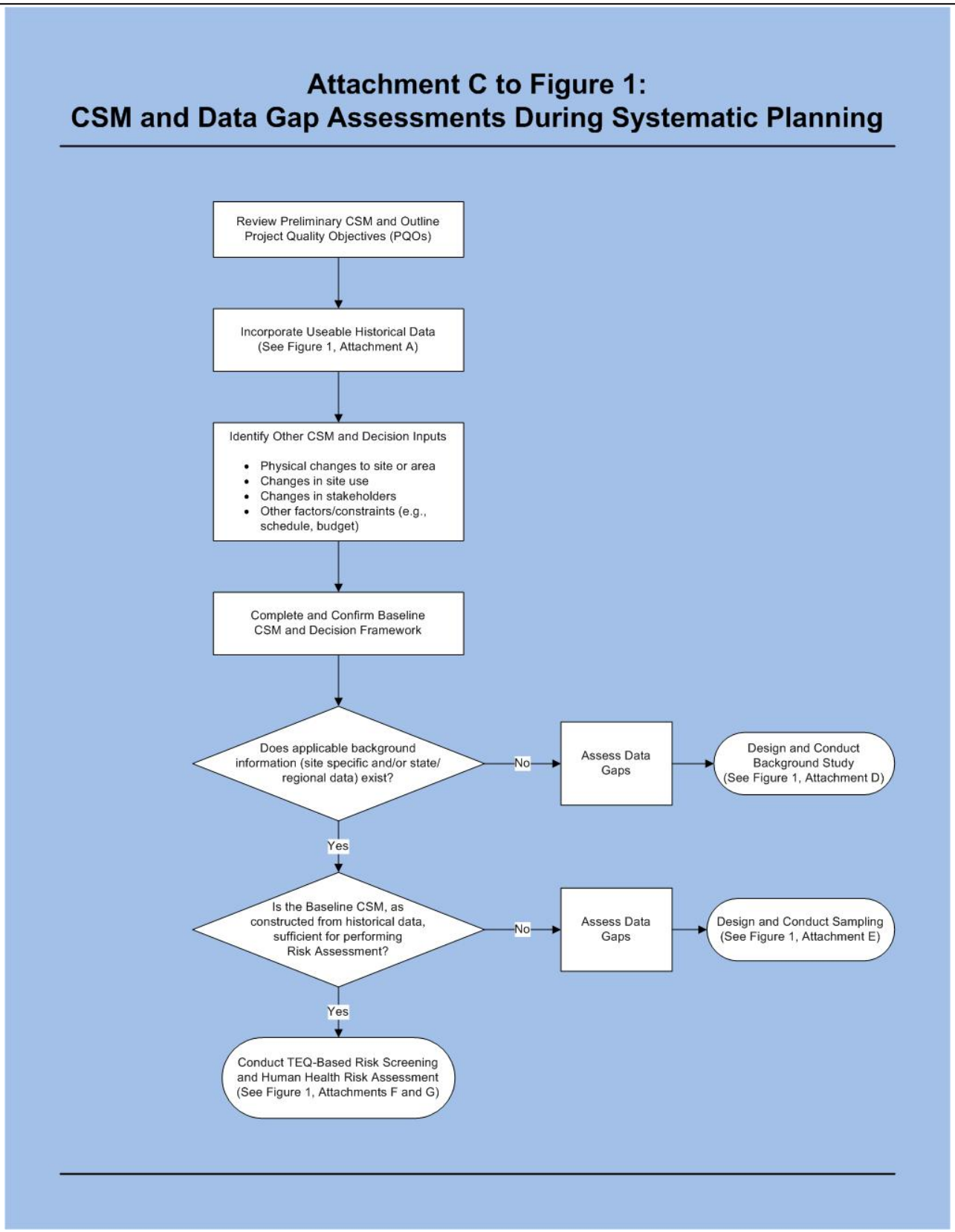
Key outputs of the systematic planning process include (but are not limited to):

- clear statements of problem definition, project goals, and the desired level of decision confidence when meeting those goals (Worksheets #10-13, 15, 28);
- sampling strategy (Worksheets #17, 18, 20-22, 26-28); and
- sampling and analytical SOPs, QC checks, performance/acceptance criteria and corrective action recommendations (Worksheets #12, 15, 16, 19 & 23-25, 28, 34-37).

- In addition to addressing scientific issues, systematic planning also considers field logistics, activity sequencing, contractual, financial/budgetary, stakeholder, social, economic, site remediation and reuse, legal, and regulatory issues (Worksheets #6, 13, 10, 11, 14, 30, 37)

Exhibits 2 and 4 provide examples of prompts for CSM inputs and considerations.

Exhibit 4: Attachment C to Figure 1 - CSM and Data Gap Assessments During Systematic Planning



Consideration of Needs for Design and Implementation of a Background Concentration Study – This section discusses considerations for whether a background concentration study may be necessary for a project of concern. A primary reason for needing a background study is to establish a reliable background concentration to comply with USEPA Superfund policy not to clean up below natural or anthropogenic background levels (USEPA, 2002d). Similarly, a background study may be needed to ensure that sampling and analysis for background samples is conducted in a manner consistent with the proposed sampling design. The determination for use of existing background information or the need for designing a site-specific background evaluation should be made by the project team during systematic planning. The rationale for whether or not a background study is required and an explanation of how background data plan to be used should be provided in Worksheets #10 and 11.

Reasons for why a background study might not be required include, but are not limited to:

- Existence of a prior background study deemed adequate during DQA efforts to be useful and relevant to site assessment;
- Existence of sufficient historical data of appropriate nature and adequate quality to assess background concentrations; or
- A determination being made that background data are not needed to manage the site.

Reasons for why a background study might be required include, but are not limited to:

- Lack of a prior background study deemed adequate during DQA efforts to be useful and relevant to site assessment; and
- No previous background data having been collected, however, systematic planning indicates the need for such information.

The process of assessing the quality and usability of the historical data (i.e., "secondary data") for the purpose of background study use should be described in the project-specific UFP-QAPP in Worksheet #13. Guidance for evaluating data is contained in *Guidance for Data Usability in Risk Assessment* (USEPA, 1991), and in Section 2.7 (page 54 of 149) of the original UFP-QAPP Manual (March 2005).

The background sampling design developed by the project team should be recorded in a "Background Concentration Study" set of sampling and analysis worksheets, including Worksheets #17, 18, 20-22, 26-28 and #12, 15, 16, 19 & 23-25, 28, 34-37. The timeline for performing the background study should be recorded on Worksheet #16. The design for comparing background and on-site concentrations should follow the guidance provided in *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (USEPA, 2002d). The statistical design should be able to meet the recommendations for statistical confidence level and power provided by that guidance.

When using an ICS design for on-site investigation, and when there is the opportunity to design the background investigation, there should be close integration of the two investigation designs. The technical elements of the background study design should mimic the on-site design. For example,

if the site DU's dimensions are 0.25 acre area to a 6-inch depth, those should be the dimensions of the background DUs. The number of increments per DU should also be the same. Soil types and particle size distributions should also be similar.

Considerations for Designing and Conducting Sampling – This section discusses considerations for designing and conducting assessment sampling. Assessment sampling design options for a site, as presented in this User Guide, are based on a unified framework developed by OSRTI in conjunction with existing EPA guidance. Final assessment designs should be developed by Regional staff on a project-specific basis and documented in the project-specific UFP-QAPP.

The data collection design is a major product of the systematic planning, as are the proposals that technical staff bring to the planning table for consideration by the stakeholders. The project-specific UFP-QAPP can be used in whole, or in part, by various project and site workers as applicable to their roles and goals. For example, the field sampling team should have ready, functional access to the relevant sample collection portions of the project-specific UFP-QAPP when working in the field. Worksheets that focus on analytical quality and instrument performance should be provided to the laboratory in advance of sampling. Options for QA include procedures to observe workers to ensure they are following the procedures described in the project-specific UFP-QAPP. The results of each assessment should be recorded on Worksheet #32.

The "Design and Conduct Sampling" section of the UFP-QAPP template provides a short summary, introducing readers of project-specific UFP-QAPPs with an overview of issues associated with the site of concern. The type of exposure scenario and type of assessment strategy the project team intends to use should be documented in applicable worksheets and attachments throughout the project-specific UFP-QAPP.

Historical data and existing information provides valuable insight to guide the requirements for assessment sampling and analysis. Even if previous data is not of the quality needed to support TEQ calculations, the data may be sufficient to help develop the Preliminary and Baseline CSMs and influence sampling density, design of DU size/shape/orientation, and locations.

Conduct TEQ-Based Risk Screening –TEQ based risk screening is a key component of any dioxin assessment. This section provides relevant information for project teams to document the TEQ screening procedures in a project specific UFP-QAPP, (Figure 1; Attachment F). Analytical results for dioxins/furans and dioxin-like PCB congeners with toxic equivalency factors (TEFs) are screened for risks to human health via the TEQ screening process, as outlined in Figure 1, Attachment F and G. (See Appendix 3 of this User Guide for the World Health Organization's 2005 TEFs.)

Calculations involve multiplying the concentration of each congener by its TEF and then summing those values to arrive at the TEQ. There are 17 dioxin/furan congeners and 12 dioxin-like PCB congeners that have TEFs and are part of the TEQ calculation. Although current analytical laboratory capability is set up to analyze for all dioxin and furan homologues and 209 PCB congeners, the OSRTI technical team is evaluating the potential to contract for modified methods that calibrate and use QC only for the 29 congeners of interest in an effort to save resources.

Human Health Risk Assessment – A Human Health Risk Assessment (HHRA) may be required based on project- or Region-specific procedures and policies. If required based on the TEQ-based PRG screening and Region-specific risk management strategies, a HHRA may be performed to determine whether detected site contaminants pose an unacceptable human health risk as related to the exposure scenarios or pathways of concern identified during systematic planning. The HHRA approach should be documented in the project-specific UFP-QAPP in Worksheets #10 and 11. The basic process for conducting the HHRA is presented in Figure 1; Attachment G.

Exhibit 5: Attachment D to Figure 1 - Background Study Design and Performance

Attachment D to Figure 1
Site Assessment Sampling Design and Strategy

BACKGROUND STUDY DESIGN AND PERFORMANCE

The determination for use of existing background information or the need for designing a site-specific background evaluation should be made by the project team during systematic planning. The rationale for whether or not a background study is required and an explanation of how background data are expected to be collected and subsequently evaluated should be provided in the project-specific UFP-QAPP Worksheets #10 and 11.

To assist project teams evaluating the necessity and design of potential background studies, this section provides several examples and considerations for collecting and evaluating background data sets in the context of proposed ICS sampling schemes. Examples provided may be sufficient to meet project objectives for some dioxin assessment projects, however site and/or decision complexity may warrant a more rigorous statistical evaluation of background. In these cases, project teams are encouraged to refer to EPA staff or technical team members with sufficient knowledge of statistics to design and implement an appropriate strategy.

OBJECTIVES AND MAIN COMPONENTS OF ATTACHMENT D TO FIGURE 1

Provide brief introduction, including:
- a) role of background evaluations in environmental risk assessment studies, and
- b) provide an operational definition for background chemicals/locations (USEPA 2002a)

Provide guidelines/recommendations for:
- a) selecting candidate background area(s)
- b) development of background sampling designs
- c) technical approaches for conducting site data versus background comparisons

NOTE: In a site-specific UFP-QAPP developed by a Region, Attachment D would be a summary description with key details of the actual background study plan to be implemented in support of the site.

CONSIDERATION FOR SELECTED STUDY COMPONENTS

A. Overall Scope and Implementation of Attachment D

In the project flow diagram (Figure 1), a trigger for Attachment D (i.e. necessity of a background study) is the absence of an existing site-specific or regional background data set (or threshold values, in the case of regional data), or a determination during review of historical documentation and data (UFP-QAPP Figure 1, Attachment A) that an existing background data set is inadequate to support site decision-making. Several examples are provided below for consideration.

Example 1. A site-specific background evaluation is recommended in cases where an existing data set for assessing background is absent.

Example 2. A tiered approach is used including an initial comparison to regional background values during systematic planning efforts (See UFP-QAPP Figure 1, Attachment C). That is, an initial screen would identify cases where site concentrations are clearly above (or below) an estimated background level. Screening against regional background values (using the highest values, if data are available from multiple background studies) would reduce the frequency of Type I (false positive) decision errors, but at the expense of inflating the Type II (false negative) error rate. This would not, however, obviate the possible need for site-specific background values further along in the process for projects entering the remediation/mitigation phase (e.g., site-specific background values would be needed if site concentrations exceed background, and background concentrations exceed the risk-based cleanup level).

B. Selection of Candidate Background/Reference Areas

Candidate background areas should be located within or proximate to assessment sites. Unlike background studies for metals, which can be ubiquitous as naturally occurring minerals, the soil type(s) and underlying geology do not have to mirror conditions present at the site (notwithstanding any issues connected with the matrix, etc. that could bias laboratory analysis) but should be matched to the extent possible to site geologic conditions. The background areas should include locations that capture local (regional) background influences, but that are not believed to have received dioxin inputs from site related activities.

Failure to include locations/areas that represent the range of potential background influences can result in biased estimates and could undermine the utility of investing in a background evaluation.

In an effort to capture the full range of potential background conditions, multiple or discontinuous background areas should be considered. Inclusion of multiple background areas may involve some form of stratified sampling (e.g., proportional allocation of sampling effort based on aerial extent of individual background areas or possibly allocation based on contaminant patterns and expected variability within individual areas).

C. Background Sampling Design

It is recommended that elements of the ICS sampling design for the planned assessment be leveraged to create a companion sampling approach for collection of the background data. As recommended for the site designs, software tools, such as Visual Sampling Plan (VSP) (Matzke et al. 2007), can be used to generate systematic random grids for the background sampling.

Background Sampling Design Considerations:

Establishing the Minimum Sample Size.

The sampling design options for assessment sites outlined in this User Guide yield a minimum of one ICS sample per decision unit (DU). There are also provisions for Regions and site-specific technical teams to collect ICS replicates for a minimum number of DUs (10% or as determined during systematic planning), another frequency of DUs, or all DUs as project data needs are determined. In cases where ICS replicates are only collected at the frequency described in the site-specific UFP-QAPP, some nominal level of replication is typically needed for the background sampling to bound the uncertainty on the background side of the site-to-background comparison. Several examples are provided below for consideration.

Example 1. Require a minimum of three independent ICS samples (1 primary and two field replicates) from one or more background areas. Use of more than one area is suggested if background is expected to vary significantly across a site, however if more than one reference area is used, then the design should be stratified to ensure background sample numbers allocated to individual areas correspond to appropriate area fractions. Individual sites would need to develop an appropriate scheme for allocation sampling effort in cases where multiple background areas are available, and there is potential for significant among-area heterogeneity in dioxin concentrations, or compositional differences in the mixture of dioxin constituents. Existing soil sampling guidance (USEPA 1990, 1992, 1994, and 2002b) adequately cover this topic and can aid project technical teams in making this decision.

Example 2. A more rigorous statistical approach is used to estimate minimum sampling parameters based on estimation or specification of desired Type I and II decision errors. Modules for treating multi-increment designs in VSP (e.g., comparing site average to a fixed threshold, comparing site average to a reference average, constructing confidence limits on a mean) allow for estimation of the number of increments per composite, as well as the number of composites, and account for blending variance and other uncertainty components. This is an option for sites with sufficient resources and technical staff, where the benefits (improved power and control of decision errors) of additional investment in the sampling design are needed to improve decision-making and can be technically justified.

D. Technical Approach for Conducting Site versus Background Comparisons

The preferred (and generally accepted standard) approach for discrete designs is comparison of the site and background distributions using two-population tests of location (typically, measures of central tendency and upper quartiles). In situations where only a small number of site results are available, an alternative is to compare individual site results to a fixed background threshold value (BTV) (See EPA [2009a]). BTVs typically represent an upper plausible limit for the unknown background population, and are estimated using simple upper percentiles of the sample distribution, or probabilistic estimators, such as upper tolerance limits (UTL) or upper prediction limits (UPL). UPLs are the preferred estimators in EPA's ProUCL software (USEPA 2009a) as well as the unified RCRA groundwater statistical guidance (USEPA 2009d).

The sample size for the site data is a constraining factor for selecting the most appropriate background screening option. Potential approaches are provided for cases where there are fewer than three, or three or more, ICS samples per site.

1. Fewer than three samples per site. This approach would compare individual ICS sample TEQ values from a site to a BTV developed for the background data set.

Options for the BTV: Some additional study and consultation with statisticians may be warranted to select either a single BTV method, or to identify criteria for selection of site-specific methods. Both the UTL and UPL should be considered as possible candidates for the BTV. Documentation of recommendations from the team statistician or statistical support personnel is suggested regarding proposed individual methods (e.g. specifying coverage and confidence level in the case of UTLs), based on the relative performance of tests conducted using these estimators for background.

2. Three or more samples per site. For cases where there is a minimum of three independent ICS samples (1 primary and two field replicates) results for both the site and background data sets, then two-population comparisons using Students t-test are recommended. This would be ill-advised for discrete designs with n=3, but for 30 to 60 increment composite samples, for both the distribution of increments within composites and the distribution of means (composites), the assumption of normality can be justified. Further, compositing dampens the variance, thus yielding relatively low estimates for the standard error. The standard error governs the effect size or minimum difference between the two means that are declared statistically significant by the test.

A two-population test for comparing the site and background means requires selection of an appropriate form for conducting one-sided hypothesis tests. EPA (2002) discusses two background tests forms. Selection of an appropriate test form has important implications in terms of pre-defining or achieving targets for Type I and II decision errors. EPA (2002a) provides additional discussion and recommendations for making this determination.

Option 1. Test Form 1 (presumption of innocence):

H0: Site < Background
HA: Site > Background

Option 2. Test Form 2 (requires a preponderance of evidence to demonstrate that the site is below background):

H0: Site > Background
HA: Site < Background

Note: Test Form 2 recommends specification of “S”, or some magnitude of difference between the site and background data that is considered significant in terms of ecological or human health effects. Project technical teams should make this determination or run the risk of flagging a problem based on a statistically detectable difference that is not significant from a practical perspective.

E**.** Selection of an optimal testing approach among competing alternatives

The selection of an optimal screening approach for conducting site versus background comparisons should be based on the relative performance of different statistical testing methods, considering constraints (e.g., sample size) imposed by the site and background data.

Relevant performance criteria include:

a) the Type I or false positive error rate (probability of erroneously concluding that site concentrations exceed background for Option 1 described above), and

b) the power of the test (the probability of correctly identifying sites that exceed background for Option 1 described above, where power= 1-Type II error rate). The power of the test depends on the "effect size", or magnitude of difference between the site and background data (averages in the case of composites) declared statistically significantly different.

Exhibit 6: Attachment E to Figure 1 - Summary of Assessment Effort, Objectives and Assumptions

Attachment E to Figure 1
Site Assessment Sampling Design and Strategy

DRAFT Uniform Federal Policy (UFP) Quality Assurance Project Plan (QAPP)

UNDERLYING CONCEPTS & DEFINITIONS

In a project-specific UFP-QAPP this section is used to describe the planned sampling design for dioxin assessment projects. This User Guide presents in-depth discussion of underlying concepts and provides definitions for terms and techniques that may be employed for an ICS design in the context of dioxin assessment. Readers are referred to the section entitled Underlying Concepts and Definitions on Page 26 of this document for concepts and definitions that can be leveraged and developed for site-specific applications.

Exhibit 7: Attachment F to Figure 1 - TEQ-Based Risk Screening

Attachment F to Figure 1
Site Assessment Sampling Design and Strategy

DRAFT Uniform Federal Policy (UFP) Quality Assurance Project Plan (QAPP)

TEQ-Based Risk Screening

In accordance with Figure 1, the analytical results from soil samples collected at sites are used to conduct a 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) TEQ-based risk screening. The TEQ-based risk screening represents a first step in risk management and Regions have the discretion to conduct human health risk assessments (HHRA) and determine the necessity of remedial actions in accordance with EPA and Regional policies and procedures. The following is a summary of the TEQ-based risk screening process.

Depending on historical information and the contaminants of concern, the dioxin TEQ ICS soil samples collected at a site may be analyzed for three classes of contaminants:

- Polychlorinated dibenzo dioxins (PCDDs); TCDD is a member of this class,
- Polychlorinated dibenzo furans (PCDFs), and
- Dioxin-like polychlorinated biphenyls (PCBs) (USEPA, 1996c).

It is well established that some of the individual contaminants in these three classes cause toxicity to humans via similar mechanisms as those causing TCDD toxicity (van den Berg and others, 2006). Historical data review should allow the site investigation and risk-screening program to focus on selected constituents within these three classes and support streamlining of the sampling and analytical program. Concentrations of each of these individual contaminants are converted to a TCDD TEQ using contaminant-specific toxicity equivalency factors (TEF). TEFs represent the relative toxicity of each contaminant to TCDD. For example, a TEF of 0.1 for Chemical A means that Chemical A is approximately one-tenth as toxic as TCDD. EPA (2009) "recommends the use of the TEFs developed by the World Health Organization (WHO)" (van den Berg and others, 1998, 2006). For the purposes of conducting the TEQ-based risk screening, the 2005 WHO TEFs should be used (UNEP, 2007).

DU-specific TEQs for the Site may be calculated as follows:

- For each detected contaminant/congener with a TEF, the reported concentration (C) of the contaminant/congener is multiplied by its congener-specific TEF to generate a congener-specific toxic equivalence concentration (TEC). Then all the TECs are summed to get a sample-specific total TEQ. For DUs with replicate ICS samples, if UCLs are to be calculated, the three (or more) replicate samples' total TEQs are used to calculate the TEQ UCL for the DU.
    - $C \times TEF = TEC$
    - $\sum TEC = \text{Total TEQ}$

- As the primary basis for decisions, the sample TEQ is calculated based on detected, non-detected, and congener results with "J" qualifiers. Non-detected congeners are included in the calculations using a Kaplan-Meier (KM) approach to determining the mean or other applicable method (see Appendix 4). Results with a "R" qualifier [representing rejected results] should not be used in the calculation of the TEQ as such, but should be used in a sensitivity analysis that evaluates the importance of the rejected data [see UFP-QAPP template Worksheet #15]. These two types of calculated TEQ-based risk screening results (with and without accounting for rejected results) can be compared and contrasted for the purpose of providing risk managers with the ability to compare to decision criteria. If the congener that is rejected is important, there are various options to pursue, including reanalysis (perhaps with improved cleanup of the sample extract). A more detailed discussion of handling non-detect and rejected data is presented in Appendix 4. A Microsoft Excel spreadsheet is available that is programmed to perform these calculations.

- Because of the possibility of reanalysis, holding times for archived samples should be tracked to ensure holding times are not exceeded, however, holding times up to one year are specified by EPA methods [see UFP-QAPP template Worksheet #19].

- Depending on project-specific applications, the total TEQ or TEQ UCL is compared to the appropriate threshold value such as, the PRG for TEQ.

If method-reporting limits (RLs) impede TEQ screening evaluations for one or more sample analyses, the affected samples may be reanalyzed to assess whether the elevated reporting limits are due to laboratory or matrix issues. If reanalysis confirms matrix interferences, the laboratory should be consulted to identify and undertake corrective actions. If matrix problems cannot be corrected, the original analytical results may be subjected to statistical evaluation to assess data usability and application.

Exhibit 8: Attachment G to Figure 1 - Conducting A Human Health Risk Assessment.

Attachment G to Figure 1
Site Assessment Sampling Design and Strategy

Uniform Federal Policy (UFP) Quality Assurance Project Plan (QAPP)

Human Health Risk Assessment

Because the Human Health Risk Assessment (HHRA) process is supported by existing guidance, which is not subject to change based on the UFP QAPP template, no information is presented in this User Guide on this topic. Reviewers are referred to the HHRA guidance listed in the References section of the UFP QAPP template and in Appendix 5: References, herein.

UNDERLYING CONCEPTS & DEFINITIONS

The following section contains explanations of concepts that underlie ICS sampling and sample handling procedures as well as assumptions for use of these procedures in dioxin assessments.

1. Sites are pre-determined to require assessment based on Regional determination.

2. Media of concern is limited to shallow soil only.

3. Shallow soil: USEPA's Soil Screening Guidance User's Guide (1996) recommends the top 2 centimeters (cm) of soil be considered 'shallow'. The actual depth interval of interest to project decisions, however, needs to be determined by the Regional project team during systematic planning according to Regional specifications or the CSM for deposition, transport and exposure. The UFP-QAPP template assumes the actual depth interval is located within the 0- to 2-foot depth interval. The chosen depth should be specified in the project-specific UFP-QAPP in Worksheet #10 as part of the "Problem Definition."

4. Deep Soil: The requirement to collect sample soils at depth needs to be determined by the Regional project team and key stakeholders during systematic planning, according to Regional specifications or the CSM for deposition, transport and exposure. The actual depth interval is assumed to be located within the 2- to 10-foot depth range. It is important to note, however, that the current version of the UFP-QAPP template does not address sampling of soils at depths greater than 2.0 feet.

5. Particle size of interest refers to the soil particle size that is of interest to the decision-making process. More information is provided in the "Matrix Heterogeneity" section of this User Guide for why particle size is important to consider. The particle size for soil is considered to be grains less than 2 millimeters (mm) in diameter. Particle size is most relevant when exploring mechanisms and pathways by which receptors can be exposed, such as dust-sized particles that easily adhere to skin and clothing. If dust-size particles are the targeted population of interest, this population can be isolated from the bulk soil by sieving. Under other exposure scenarios, it is possible that the entire particle size distribution is the parameter of interest, in which case the bulk soil (all material less than 2 mm) would be the targeted population. The particle size(s) of interest should be determined during systematic planning and in consultation with the project risk assessor.

   Standard tests of soil properties, especially particle-size analysis, should be performed on at least one soil sample collected from the site. If the site encompasses significantly different soil types, then areas composed of each soil type should be sampled and tested. If a background concentration study is performed, particle-size analysis can be performed on a representative sample during this step. As an added benefit, particle-size information can serve multiple uses over the course of site characterization and remediation.

6. Targeted population refers to the specific material or group of objects that are of interest to an investigation. For the purposes of the UFP-QAPP template, a target population is

the soil area, soil depth, and particle fraction about which project decisions are made. Defining the target population is one of the most important outcomes of the systematic planning process.

7. Sample support: "Support" refers to the physical parameters that help determine contaminant concentration results when a sample is analyzed. "Sample support" includes the mass or volume of a sample, along with its dimensions (length, height, width) when extracted from its parent matrix. When soil is involved, "sample support" also includes particle size, as this has a marked influence on the mass of contaminant attached to a mass of soil. The term "support" also conveys the spatial dimensions and other physical characteristics of the population of interest. "Decision unit support" is simply a shorthand way of saying "the spatial and physical dimensions of the decision unit." For example, a DU support might be the particle size fraction less than 100-mesh that is within the top 6 inches of a quarter-acre DU.

   EPA's Soil Screening Guidance User's Guide (1996) notes that the size, shape, and orientation of sampling volume (i.e., support) for heterogeneous media have a significant effect on reported measurement values. For instance, particle size has a varying effect on the transport and fate of contaminants in the environment as well as on the potential receptors. Comparison of data from methods based on different sample supports can be difficult. Defining the sampling support is important in the early stages of site characterization, and is a task inherent to the "set the study boundaries" DQO process step.

8. An ICS sample is defined as a single sample composed of soil portions (increments) collected from multiple locations evenly distributed across the designated volume of a designated area. The soil increments should all be of equal mass and mixed together to form a single large uniform sample. Use of ICS can require additional sample preparation/homogenization before being sub-sampled to isolate the portion of the sample volume submitted for analysis. The concentration of an ICS sample represents the *average concentration* for the soil volume of the designated area. The designated volume of soil represented by the ICS sample may be either a DU or an SU, depending on the specific structure of the particular sampling design and the purpose of the ICS sample.

9. Compositing vs. incremental sampling vs. incremental compositing: All EPA guidance on compositing was published prior to the term "incremental sampling" coming into vogue in recent years. EPA guidance uses the term "compositing" to cover a variety of different sampling goals: 1) deriving an average, 2) searching for hotspots, and 3) estimating a population proportion (USEPA, 2002). Recently, the term "incremental sampling" has come into use only in the context of deriving an average. A number of organizations are still in the process of defining what characteristics should be present in order to use the term "incremental sampling." Both compositing and incremental sampling refer to the same basic process of mixing portions together; however, it is the purpose and details that differ. EPA's "compositing" term is broader in its goals and applications than incremental sampling. For the purpose of the UFP-QAPP template, incremental sampling is considered a specific type of compositing used to derive an

average. In some instances, dioxin sampling designs may also involve searching for hotspots. Since one or both of these goals might be present in a single sampling design, the term "incremental composite sampling" (ICS) is used herein to cover all possible sampling goals.

10. Application of composite or incremental sampling in USEPA's Superfund program:. USEPA's 1996 *Soil Screening Guidance User's Guide* reasons: "Because the objective of surface soil screening is to estimate the mean contaminant concentration, the physical "averaging" that occurs during compositing is consistent with the intended use of the data. Compositing allows sampling of a larger number of locations while controlling analytical costs…" (USEPA, 1996a).

    The 2002 *Supplemental Guidance for Developing Soil Screening Levels for Superfund* presents composite sampling as an alternative to collecting the large number of discrete samples that would otherwise be required to generate an estimate of the true mean. The 95% upper confidence limit (UCL) was introduced as a technique to compensate for the inaccuracies of estimating the mean from a small number of discrete samples, as is generally done, even when spatial variability is high:

    > "The maximum contaminant concentration from composite sampling is a conservative estimate of the mean concentration and can be used for soil screening evaluations… Alternatively, site managers can collect discrete un-composited samples…[but] because there is no spatial averaging of soil concentrations with this method, a much larger number of soil samples is required to produce a reliable estimate of the mean contaminant concentration. As a result, EPA recommends estimating the 95th percentile upper confidence limit (UCL_{95}) on the mean contaminant concentration as a conservative estimate of the mean when performing a soil screening evaluation with data sets of non-composited samples."

    (See also the 1996 *Soil Screening Guidance Technical Background Document*.) The use of the UCL95 is one alternative that should be considered during systematic planning. Regions and project teams have the discretion to use other threshold values (USEPA, 1992a).

    The USEPA *Risk Assessment Guidance for Superfund* (*RAGS*, 1989c), says that composite samples are acceptable for representing average exposure concentrations:

    > "For media such as soil, sediments, and groundwater, composite samples generally may be used to assess the presence or absence of contamination; however, they *may be used in risk assessment only to represent average concentrations (and thus exposures) at a site*." [page 4-19; emphasis added].

    At least six EPA guidance documents address compositing, collectively providing many details for how to use compositing for different project purposes. In addition, advances in technology have deepened the understanding of natural systems and resulted in the development of more powerful field and laboratory tools. These tools present RPMs with

even more options for planning and implementing compositing projects than were available when these guidance documents were originally published. EPA guidance and findings from current (2010-2011) research on incremental sampling and sample homogenization were used to prepare the UFP-QAPP template and this User Guide.

11. A decision unit (DU) is defined herein as the volume of soil over which a mean concentration value is obtained for comparison to a regulatory threshold value or other type of action level or for using in risk assessment calculations. It is the same as the material within the "study boundaries" as covered in Step 4 of USEPA's DQO process. The study boundaries "specify the spatial and temporal aspects of the environmental media that the data must represent to support the decision." (USEPA, 2006).

    In risk assessment, the average concentration obtained over a DU is called the exposure unit (EU), and is termed the exposure point concentration (EPC). Thus, the DU is the area and volume about which the primary project decisions (such as evaluating risk or the need for remediation) are made. The 1996 SSG User Guide displays a graphic on its page 11 (http://www.epa.gov/superfund/health/conmedia/soil/index.htm) showing how a site can be stratified into different sections based on their likelihood of contamination. Later in this dioxin assessment User Guide, Example 2, beginning on page 56 is loosely based on the SSG graphic and illustrates its application.

    The average concentration for a DU can be derived by:

    - Collecting multiple increments and analyzing a single ICS sample that provides an average for the whole DU. This is illustrated in Figure 1 below. A UCL may also be calculated by using an estimate of variability from a similar (same land use, physical attributes, etc.) nearby DU as described in Example #2 in the implementation section of this User Guide.

    - Collecting multiple ICS sample replicates (i.e., take a similar set of 3 or more samples from the same DU, but offset the grid so that increments are from different locations), analyzing ICS samples for the DU and calculating a UCL from replicate results. This is illustrated in Figure 2 below.

If the primary DU ICS sample fails and the sampling design includes SUs designed to assist potential remediation planning, averaging analytical results from multiple SUs within the DU and calculating a UCL from that data set is an option. This may occur in a scenario where a hotspot is identified in a single SU, which is then remediated. If an updated UCL is required for the DU, the remediated SU can be sampled using the same number of increments and ICS samples as the pre-remediation SU sample. The post remediation SU results can be substituted for the pre-remediation SU results. The new data set (original "clean" SUs and the post-remediation SU sample results) can be used to calculate a post remediation UCL for decision making. Note that post remediation results are still based on the same number of increments as the pre-remediation DU (greater than 30 increments). Figures 1 and 2 below illustrate the collection of increments using a random systematic grid and highlight the need for off set grids to collect ICS replicates within the same DU.

Figure 1. Single DU with 30 Increments Going into a Single ICS Sample

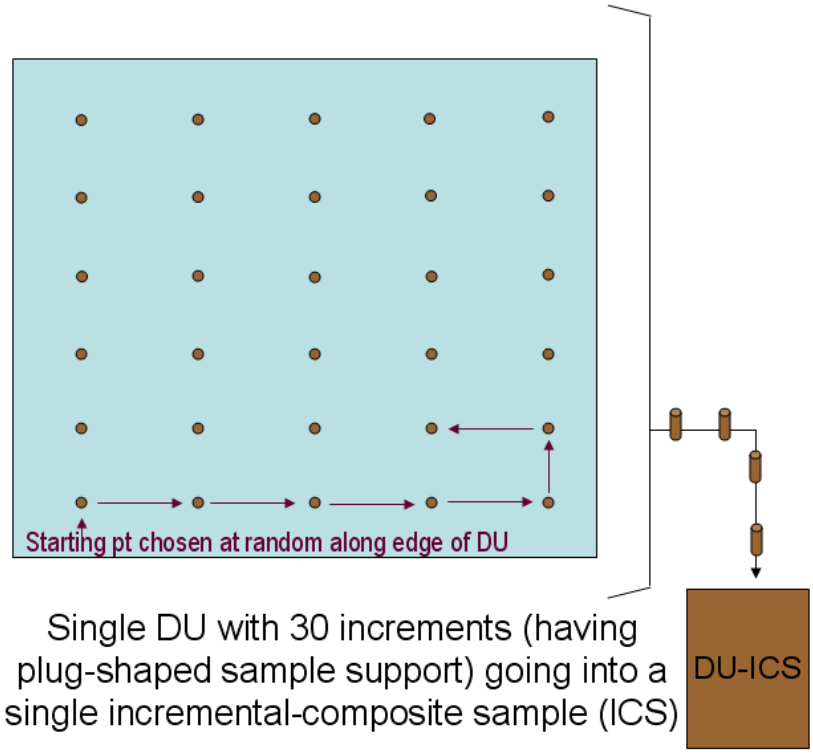
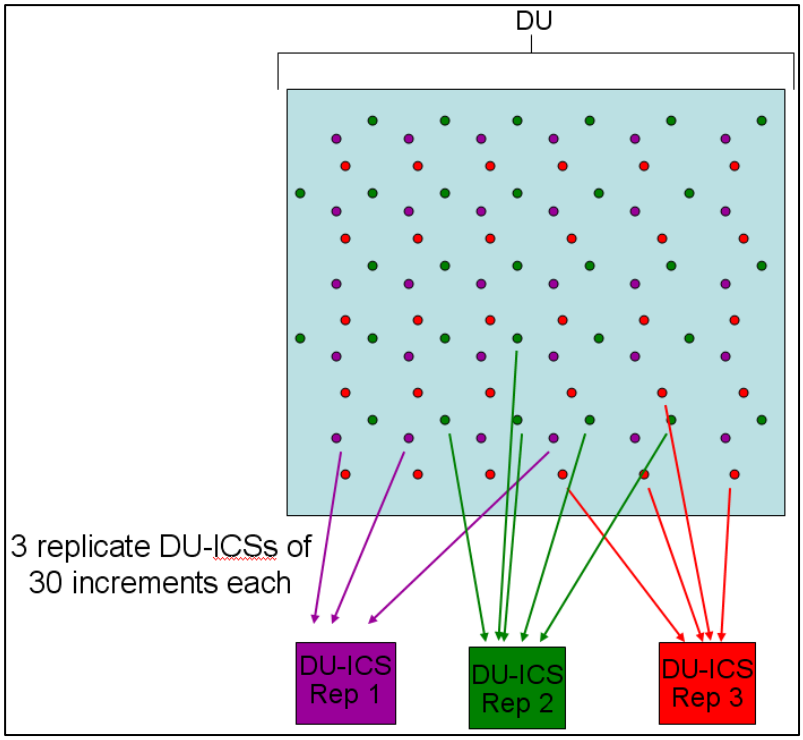


Figure 2. Three Replicate DU-ICS Samples of 30 Increments Each



12. Exposure point concentration: EPA recommends that comparisons with risk-based guideline values and standard risk equations use a concentration statistic that represents the average exposure of a chemical that a receptor may be exposed to within a given area. EPA's *RAGS* (2002) states that:

    > "the exposure point concentration (EPC) is a conservative estimate of the average chemical concentration in an environmental medium. EPA recommends using the average concentration to represent 'a reasonable estimate of the concentration likely to be contacted over time."

    The conservative estimate of the average recommended by EPA is the statistical upper confidence limit on the mean. This agrees with the *USEPA's Soil Screening Guidance User's Guide* (1996a):

    > "an individual is assumed to move randomly across an exposure area (EA) [note: this is the same as the current term "exposure unit"] over time, spending equivalent amounts of time in each location. Thus, the concentration contacted over time is represented best by the spatially averaged concentration over the EA. Ideally, the surface soil sampling strategy would determine the true population mean of contaminant concentrations in an EA. Determination of the "true" mean would require extensive sampling at high costs; as a result, the maximum contaminant concentration from composite samples is used as a conservative estimate of the mean. The Max test strategy compares the results of composite samples to the SSLs [soil screening levels] of 1996."

    Pages 13 and 14 of the *Soil Screening Guidance* User's Guide (1996a), describe a compositing strategy, including selecting the number of increments based on knowledge or assumption of the amount of variability within the EA. This example is similar to the strategy as recommended herein, with minor differences such as using a random placement of increments rather than random-start gridding as described herein.

    If it is acceptable to compare composite samples to soil screening levels (SSLs), what does this imply about compositing for a PRG? The *Supplemental Guidance for Developing Soil Screening Levels for Superfund Sites* (2002) states:

    > "EPA recognizes, however, that certain conservative assumptions built into the generic and simple site-specific approaches to SSL development, while appropriate for a screening analysis, may be overly conservative for setting PRGs…such conservatism may not be necessary for developing PRGs and cleanup levels for many contaminants."

    This language implies compositing is considered appropriate for a conservative screening technique, leading to the conclusion that it would also be appropriate for another similarly conservative decision goal.

13. An exposure unit (EU) is the area over which the EPC is calculated. It is the same concept as the EA described above, although EAs in the Soil Screening Guidance are presented as 0.5 acre. For residential areas subject to dioxin site assessment, a 0.25-acre EU is recommended as a default. The size of the DU may, however, vary from 0.25-acre depending on risk assessor needs, site-specific parameters, land use, and other driving factors of the project. The basis for EU sizing should be outlined in the project-specific UFP-QAPP in Worksheet #11. The EU is a common type of DU, and consistency with the term "unit" is a reason this User Guide uses the term "exposure unit," rather than "exposure area." In addition, by 2001, EPA RAGS guidance was using the term "exposure unit."

14. A sampling unit (SU) is equivalent to an "area of inference," a term used in EPA's PCB guidance (1998) where compositing is recommended for PCB characterization (http://www.epa.gov/wastes/hazard/tsd/pcbs/pubs/subpartmopr.pdf). A SU is a volume of soil represented by a single sample, which potentially can be analyzed for a data result. That single result might be from a single discrete sample in the center of a grid cell (the cell being the SU), or, as in an ICS design, the single data result might come from an ICS sample from across the grid cell. In other words, instead of trying to represent the concentration of the whole grid cell from a single discrete sample, the grid cell is represented by a single sample, which is made up of many increments collected from across the grid cell.

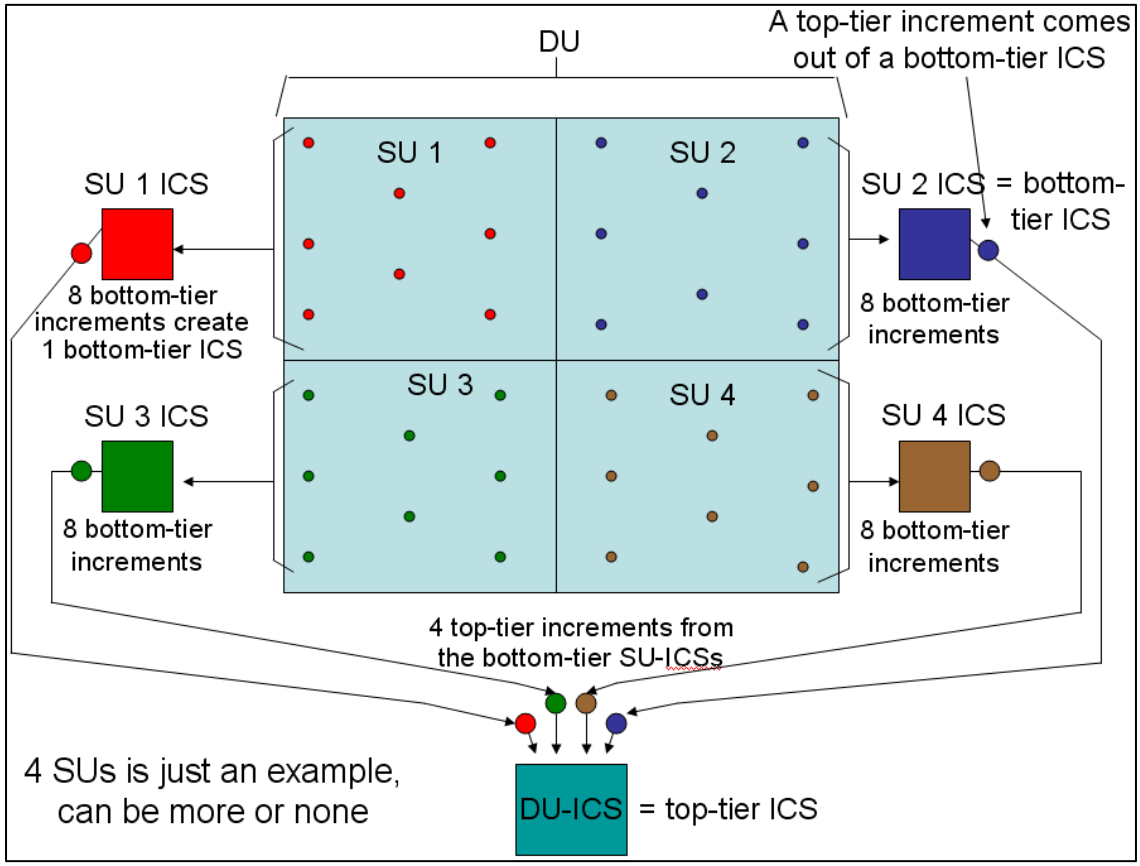
    If an ICS sample is composed of increments taken from across the entire DU, the ICS sample represents the entire DU. In that case, the DU is also a SU (because the DU is represented by only a single sample). However, a DU can be partitioned into subunits (i.e. SUs), each of which is sampled with its own composite sample. Such a subunit is a SU, because each subunit is represented by a single sample. Each SU composite sample (abbreviated as SU-ICS) can be separately analyzed as individual samples. More likely however, each SU-ICS sample should be split or sub-sampled. Then the splits would be composited together to create the DU-ICS sample that represents the entire DU. See Figure 3 below, where a DU is divided into SUs as quadrants. Increments are collected from each quadrant-SU to represent the whole SU's volume. Compositing of the SUs then produces a single ICS sample that represents the whole DU.

    It is easiest if all SUs within a single DU that are sub-sampled to create the "top-tier" DU ICS sample are of the same size area (see the figure below) so that each SU contributes equally to the DU composite. Otherwise, the amount of soil contributing to the DU's ICS sample would need to be weighted to reflect the proportional area of each SU. Depending on the sampling design and within SU variability, SUs do not always need to have the exact same number of increments. The goal is for the set of SU increments to represent the average concentration for the SU. Unless the concentration variability across an SU is known, the exact number of increments needed (as determined statistically) to represent the average for that SU cannot be known. However, in most cases, the spatial density of increments (i.e., the number of increments per unit area) will be approximately the same for all the SUs within a DU, and therefore can be assumed to reasonably represent the mean for each SU. Where this might not hold true is when

different SUs have markedly different concentrations, such as when one SU contains a "hotspot." If the variability is high within the contaminated SU, the number of increments might not be sufficient for an accurate estimate of the true mean of that SU. However, in that case, if the concentration is well above an action level, it is not as critical that the mean estimate be as accurate as when the concentration is nearer an action level. Where a more accurate estimate of the mean is required for each SU and variability is expected to be high, 30 increments or more per SU may be used for the reasons explained in Appendix 2. This is more likely to apply at sites with larger DUs.

The use of SUs makes the sampling design somewhat more complex, and the UFP-QAPP template anticipates that most of the time the SU concept may not be utilized. At times however, there might be site-specific reasons to split the DU into SUs. These include the need to document spatial trends in concentration within the DU if a CSM has data gaps that need to be filled at a sub-DU scale. In addition, the project team may suspect that contamination is not uniformly distributed within the DU (i.e., "hotspots" might be present), or that exposure to one SU (such as a child's play area) is higher than to other SUs. SU samples can be archived for later analysis in case the DU fails and is determined to require cleanup or further investigation. Archiving SU samples for later analysis helps to reduce the chance for multiple field mobilizations, and their associated work plans, sample acquisition, analysis, data interpretation and reporting.

Figure 3: Example of Four SUs Within a DU



Compositing SU ICS samples to create a DU-ICS sample is the physical equivalent of taking four discrete samples (one from each quadrant) and averaging them together, but with two very important differences. First, the uncertainty about whether a single grab sample is representative of its quadrant is controlled by using a composite sample (made of bottom-tier increments) rather than a simple discrete sample. In this way, the SU increment density is serving to control the short-scale heterogeneity over the quadrant, ensuring the representativeness of the SU ICS sample. The SU ICS sample is assumed to *approximately* represent the average concentration over the SU. The second difference is that, rather than analyze each quadrant separately and then mathematically calculating an average for the entire DU, this is done physically through a compositing process. The remainder of the SU ICS sample can be archived for future use, if needed.

In the end, the number of increments contributing to the DU-ICS sample is 30 to 60 (unless substantially justified in the project-specific UFP-QAPP), a number of increments that provides more confidence that the result represents the true mean, which is critical for risk-related decision-making on the DU. However, if the DU is found to be contaminated at a concentration level where remediation may be necessary, the SU ICS samples can be retrieved from archives and tested to determine the contaminant pattern at a spatial scale smaller than the DU where such information might inform remedial decisions.

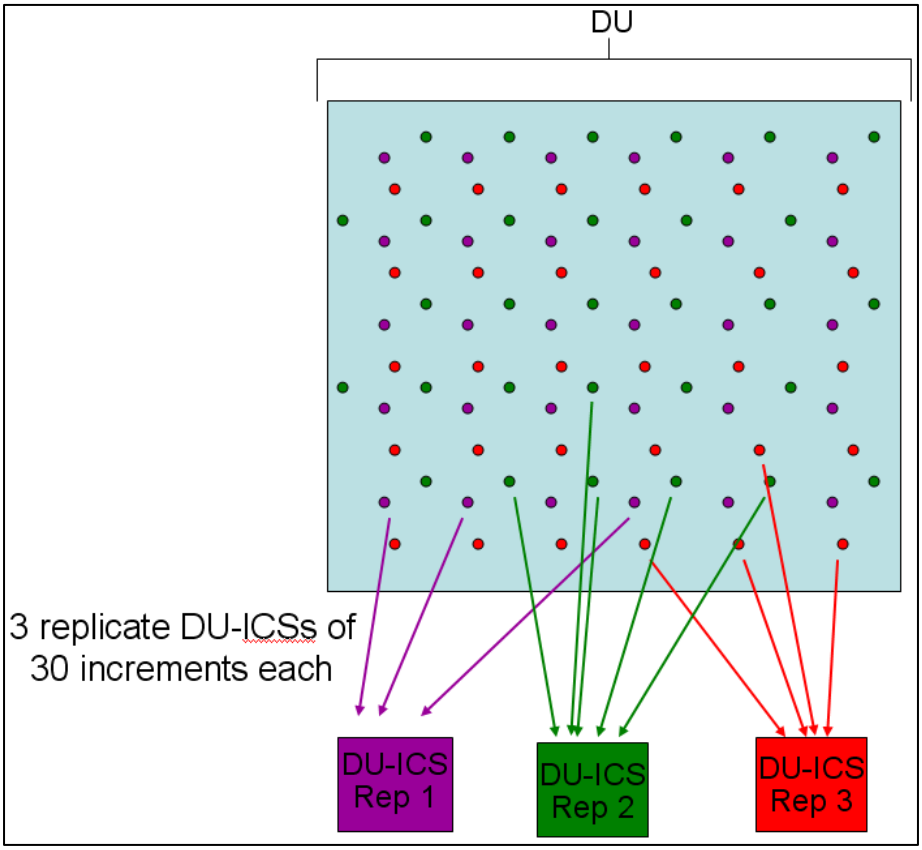
15. When both DUs and SUs are used, terminology should distinguish between the ICS samples produced at the SU level, and the ICS samples produced at the DU level. Shorthand designations such as “SU-ICS” and “DU-ICS” might be used. There may also be times when terminology such as “bottom-tier” (i.e., at the SU-level) and “top-tier” (i.e., ICS samples formed by combining other homogenized ICS samples) might be useful. Any use of tiered ICS samples should be clearly described as part of the sampling design in the project-specific UFP-QAPP in Worksheet #17.

16. Collect samples for other analyte species if planned, including discrete samples for analytes requiring undisturbed samples (such as volatile organic compound (VOC) samples) or other project data needs. It is also possible to collect VOCs using a compositing technique that places increments directly into methanol preservative (USEPA SW-846 method 5035) http://www.epa.gov/epawaste/hazard/testmethods/sw846/pdfs/5035.pdf

17. Ideally, the number of increments should be selected based on the degree of heterogeneity in the field and the variability seen in sub-sampling. If sufficient historical and other information is available, it is possible to use Visual Sample Plan (VSP) to calculate the ideal number of increments. This information may also be obtained from a site-specific pilot study or demonstration of method applicability (see Appendix 3). For those cases where neither option is available, a default number of 30 to 60 increments per DU ICS sample is proposed (see Appendix 2). This number is considered the default, meaning that if 30 to 60 increments per DU ICS samples are collected, no explanation of how the number was selected is required. One of the reasons a number higher than 30 might be utilized is if the CSM indicates high heterogeneity, but a quantitative measure is not available. If a TEQ UCL (an upper confidence limit on the mean) is desired, rather than the simple TEQ mean (as represented by a single ICS sample result), then three replicate

ICS samples can be collected, tripling the spatial coverage of the DU (see figure below). When replicates and a UCL are used, a number closer to 30 may be sufficient for even a highly heterogeneous DU, because a total of 90 increments (three x 30-increment ICS samples) is collected within that DU. If replicates are not used (only a single ICS sample per DU), then project teams may choose a number closer to 60 increments for the single ICS sample in order to ensure an accurate estimate of the mean for DUs suspected of having a high degree of heterogeneity.

Figure 2: Three Replicates DU-ICS Samples of 30 Increments Each. (Figure Duplicated here for Reference)



The number of increments can be optimized to more than 60 or less than 30 if the reason for the deviation is scientifically- or statistically-derived, is transparent, and the reasons are clearly explained in the project-specific UFP-QAPP in Worksheet #17. Referencing guidance that deals with a different analyte is not considered adequate justification.

The default minimum increment number of 30 was selected based on a number of factors, which are discussed in detail in Appendix 2, "Rationale for a Default of 30 – 60 Increments."

Calculating UCLs from ICS replicates

One of the reasons a default minimum of 30 increments was selected is that the Central Limit Theorem and its application in research studies suggest that this number is sufficient to normalize non-normal populations. Stated more rigorously, repeated sampling of a non-normal population typically generates a distribution of means that is normal or near-normal. As long as there is no evidence to negate the assumption of near-normality, use of t-distribution statistics for calculating the UCL is allowed. Being able to use the t-distribution simplifies the mathematics and generates a "well-behaved" UCL. However, if there is evidence that the assumption is not true for a certain DU, then an alternate statistical method to calculate the UCL is required.

A group of statistician members of the Interstate Technology and Regulatory Council's (ITRC) Incremental Sampling Methodology (ISM) Team recently performed simulations exploring this question. Their simulations were based on using three replicate incremental samples per DU. The following recommendations are based on their conclusions.

- Evidence that the distribution may be non-normal can be extracted from the three replicate data points by calculating the relative standard deviation (RSD), which is the standard deviation divided by the mean. A high RSD is an indication that the distribution may be significantly non-normal. For that reason, using the t-distribution to calculate the 95% UCL is recommended only when the RSD is less than 1.5.
- For RSDs between 1.5 and 3, their recommendation is to use the nonparametric 95% Chebychev UCL.
- For RSDs greater than 3, use of the nonparametric 99% (not 95%) Chebychev UCL is recommended.

Note that ProUCL , an EPA software tool that can calculate UCLs for a variety of distributions (*http://www.epa.gov/esd/tsc/software.htm*)will not calculate a UCL for data sets of only three samples because the algorithms used cannot make the choice of which UCL to recommend. A spreadsheet programmed to calculate these values from data results, therefore, has been developed for use in conjunction with the UFP-QAPP template. Other commercially available software, or calculation by hand may also be used by qualified technical team members.

The UCL from the Chebychev calculation, especially the 99% Chebychev UCL, can be quite high. It is desirable, therefore, to control as many variables as possible so that the ICS sample replicate variability (the standard deviation) is low. Achieving this involves 1) using the CSM to separate different populations, 2) conducting dense sampling of the population with sufficient increments, and 3) performing careful homogenization and sub-sampling of the ICS samples. These activities help make each ICS data point highly representative of the true DU mean, and therefore make the underlying data distribution normal. With a normal or near-normal distribution, the RSD should be less than 1.5 and the t-distribution can be used to calculate the UCL.

18. Pilot studies can be very cost-effective in addressing a variety of important unknowns. A fact sheet describing the benefits of pilot studies [also called "demonstrations of methods' applicability" (DMAs), a term from SW-846] is attached as Appendix 3. One example of how a DMA can pay for itself is when there are large areas made up of many DUs and the CSM indicates that all the DUs would have a similar variability. If that variability is low, performing a DMA on a few DUs may show statistically that fewer than 30 increments per DU are acceptable. In that case, given the number of DUs involved, reducing the number of increments to only what is statistically required could result in considerable labor savings. Visual Sample Plan (VSP) can be used to perform calculations on DMA data to statistically determine the number of increments. Documentation providing the derivations of VSP inputs should be provided in the project-specific UFP-QAPP in Worksheet #17 The DMA may provide evidence that not all DUs require field replicate DU-ICS samples. Perhaps only a certain fraction need the replicates as a QC check to ensure that conditions are stable across the large area. Technical suggestions for implementing a DMA specific to dioxin assessment projects, along with an instruction sheet for using VSP in this context may be developed from early pilot sites to assist Regional project teams in future assessments. Project specific assistance is also available from EPA HQ/Superfund/TIFSD.

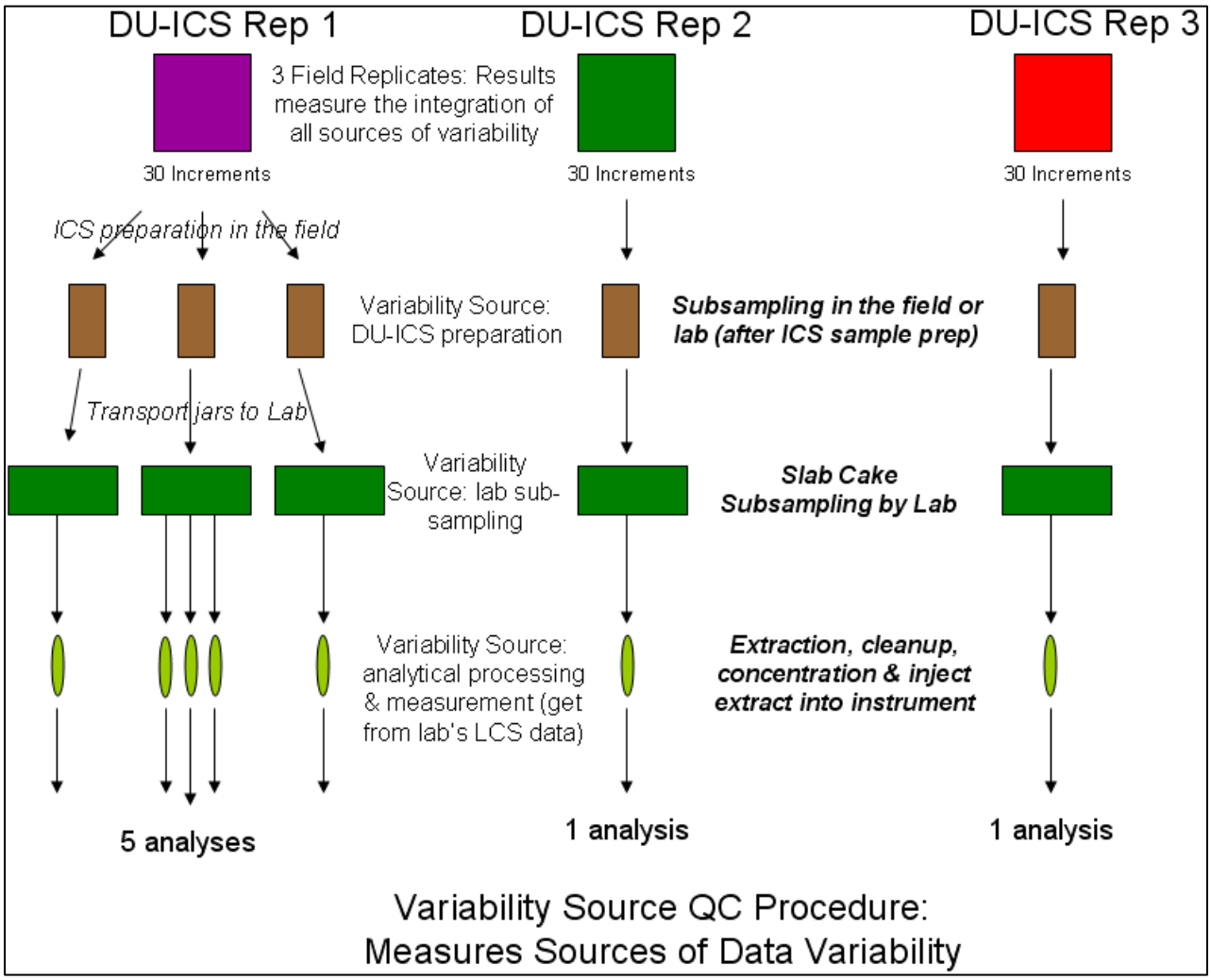
    Another reason for performing a DMA when many related DUs are involved in a single project is to make sure the sampling and analytical methods can address matrix-specific challenges and meet the project-specific goals. It is a good idea to test SOPs and ensure they are adequate for increment collection, sample preparation, and sample analysis before moving forward. Finding out too late that SOPs are inadequate can result in inefficient use of valuable resources or require repeat site activities. Similarly, standard sample preparation methods may be over-kill in some situations, and modification of the SOPs to reduce labor and associated costs may be beneficial.

    When a DMA is performed, it is a good idea to determine which step in the sampling and analysis chain presents the greatest source of data variability (USEPA, 1996a) as it is important to identify large sources of data variability that may jeopardize project goals. If any adjustments in analytical or sampling procedures are needed, it would be most efficient to determine those before project work plans are finalized. Focusing modifications on those procedures having the highest impacts on data variability or data usability prior to full-scale field activities can avoid wasted labor and unusable data.

    Figure 4 shows the construction of a variability source QC design that measures sources of data variability. Seven analyses are required for each sample evaluated, whereas, the number of individual samples evaluated is flexible depending on what else the DMA is testing. Based on the DMA results, the project team is able to confidently specify the details of sampling and analytical procedures and develop QA/QC procedures that build transparency and defensibility into the project while saving resources. An example of using such a study is illustrated in Example 2 of the implementation section of this User Guide.

Even if a DMA is not performed, the investigation should be structured to include some type of variability source QC procedure (USEPA 1996a, 1990). The figure below illustrates a procedure that measures the relative strength of variability sources. This procedure should be performed at a rate determined by systematic planning and as documented in the project-specific UFP-QAPP worksheets. When possible, the project team should target DUs that have different characteristics that could influence the efficacy of the SOPs, such as matrix properties and release and transport mechanisms.

Figure 4. Variability Source QC Procedure: Measures Sources of Data Variability



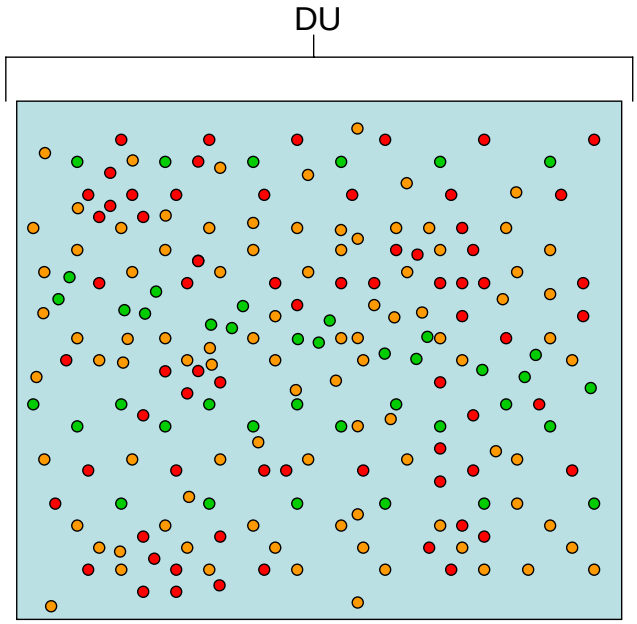
19. What if the DMA or CSM indicates that more than 60 increments are needed? For example, historical data may indicate a very high degree of variability is present. General experience is that collecting more than 60 increments per ICS sample begins to make the logistics of mixing, disaggregation, sieving, and other sample preparation procedures proportionately difficult to implement. Where statistics indicate that more than 60 increments are required, changing project plans to utilize smaller DUs may be beneficial.

Justification for increment numbers greater than 60 should be documented in the project-specific UFP-QAPP.

If a large number of increments are suggested by a statistical evaluation, the sampling and analytical design should be examined more closely in light of the CSM, and/or DMA data, and variability source QC samples' results (Figure 4). For example, the variability in the data may result from inadequate sample homogenization or poor analytical sub-sampling and not *in situ* field conditions. If the variability problem lies in sample handling, corrective action can be taken at that level. Collecting a larger number of increments is not beneficial if the greatest sources of variability are the result of laboratory sub-sampling or other handling procedures.

If the source of high data variability appears to be spatial variability, the CSM may need to be updated and the reason explored. The updated CSM may suggest that the variability is mostly uniform but randomly distributed across the DU area, as illustrated in Figure 5 below.

Figure 5. Example of Random Distribution of Contamination



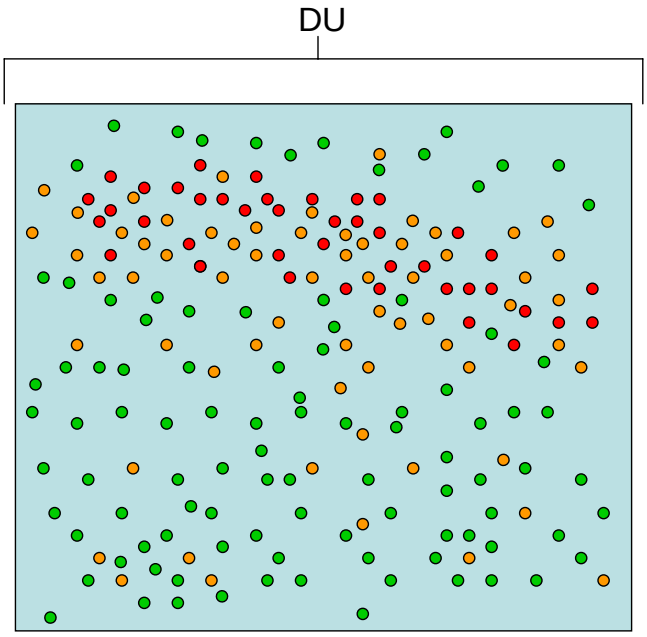
Random distribution of contamination = a single heterogeneous population of highly variable contaminant concentrations

- High contaminant concentration
- Medium cont. concentration
- Low contaminant concentration

In this case, the CSM should be assessed to determine whether contaminant release and migration created patterns of contamination distribution that are highly variable even on a

small spatial scale, or biased by a unique physiographic feature, as in the example Figure 6 below, which shows distributions based on the presence of a swale on site.

Figure 6. Example of Spatially Stratified Distribution of Contamination



Spatially stratified distribution of contamination = 2 populations of contaminant concentrations as a result of a swale

- High contaminant concentration
- Medium cont. concentration
- Low contaminant concentration

In some cases, observable DU-specific characteristics or unique physiographic features contribute to causing different "populations" of contaminated soil in different parts of the same DU. For example, consider a site where contamination is suspected to have occurred by spraying. It is possible that a DU in the investigation area was not directly sprayed, but spray drift likely deposited a lesser amount of contamination. Drift deposition could cause short-scale heterogeneity—where one potential core sample can differ from its neighbor core sample. This is illustrated in Figure 5.

On the other hand, suppose there were a swale that collected storm water runoff from the directly sprayed area, which channeled the runoff onto the DU (see Figure 6). The part of the DU that lies in the swale would likely contain higher contaminant concentrations than the rest of the DU. Thus, that swale would represent a different contaminant population than the rest of the DU because contamination migrated there in a different manner, creating a distinct spatial pattern and concentrations different from the rest of the DU. The difference from one part of the DU to the other (the swale) creates "long-scale

heterogeneity." To reiterate, long-scale heterogeneity means that samples from one location are reasonably similar to each other, but significantly different from samples from other locations in the DU.

20. When data from different contaminant populations are mixed, it creates high data variability and non-normal data distributions. Separating populations reduces data variability and makes statistical distributions more normal, which greatly simplifies statistical work and improves remedial design. Within the ICS framework, a DU known or suspected of containing different populations might be split into separate DUs that each hold just one population. For example, a wind-deposited population versus a water-deposited population versus a spillage area, etc. Often there are observable physical features or clues in the CSM that help delineate these different populations.

Matrix Heterogeneity, Decision Units, and Sampling Units

There are a few fundamental concepts behind the ICS design rationale in the UFP-QAPP template. One of the most important concepts is that soils and similar solid media can display a high degree of heterogeneity that greatly complicates collecting data representative of the population of interest. A further complication is that the degree of heterogeneity depends on the spatial scale being examined. The challenge for site characterization is that the scale of data sampling and analysis is much smaller (grams) than the scales at which exposure characterization and remediation take place (thousands of kilograms). The goal of ICS is to bring those two spatial scales into better alignment so that the data from the analysis of 1 or 10 grams of soil can be confidently extrapolated to represent the concentration of a volume of soil in the field that is many orders of magnitude larger.

In the context of project planning and implementation, defining and sampling DUs is a fundamental step. Defining appropriate size, shape, and orientation of DUs is, therefore, a primary step and critical for systematic planning for any dioxin assessment effort. In this respect, most site decisions are made at the DU level and efforts associated with managing heterogeneity or sources of variability work down to shorter scales from there. In contrast, quality data collection and management starts at the sampling or matrix level and moves up through the DU level. Project decision makers, and those charged with designing and collecting information to support those decisions, therefore, may view heterogeneity from opposite ends of the spectrum. In an effort to more fully describe techniques to manage sources of variability at differing scales, this section describes heterogeneity from the matrix micro-scale to the DU level.

Matrix heterogeneity refers to variation in composition, and especially (for the purposes of this User Guide), the variation in contaminant concentrations from place to place within soil over the volume of the DU. Variation in soil composition is correlated with variation in particle size and contaminant concentration. The term "within-sample matrix heterogeneity" refers to the variation in composition that causes variability in data results between duplicate sub-samples from the same jar. This within-sample heterogeneity acts at the micro-scale level, which involves the soil particles and contaminant molecules interacting with one another. Heterogeneity also occurs at larger scales and has an effect on data sets and their interpretation, as discussed in more detail in the following paragraphs.

Within-Sample Matrix Variability

Starting at the smallest spatial scale, data variability caused by within-sample matrix heterogeneity is the first source of variability. This source of data uncertainty is measured by laboratory and field replicates. These splits are from the same sample location and should theoretically give the "same" result, but due to matrix variability, discrete samples often do not. When sample homogenization is minimal, sample splits are not the "same" sample; as would be reflected by the different analytical results obtained. The data uncertainty being created by within-sample heterogeneity increases as the difference between replicate results grows.

Within-sample matrix heterogeneity is reduced by sample preparation activities, such as drying, disaggregation (breaking up clods) and sieving to separate particle size fractions. [Note: Grinding of a sample may be necessary for some analytes or some scenarios. Based on the experiences of practitioners with dioxin-contaminated soil, the UFP-QAPP template assumes that sample grinding will NOT routinely be required. Since that cannot be guaranteed, the grinding option has been retained, and the UFP-QAPP template developers can provide assistance should the issue arise. The experience and SOPs from such a project can then be incorporated into updates to the UFP-QAPP template and distributed to other Regions.]

Matrix heterogeneity at the micro-scale is a function of soil particle size and particle composition. These soil particle properties cause preferential retention of analytes on certain soil particles rather than on others. Soil properties such as surface area and electrostatic charge can affect the degree to which some analytes of interest preferentially sorb to those materials. For example, clay minerals carry an electric charge that attracts positively charged metal ions. Similarly, particles composed of significant amounts of organic carbon may accumulate higher levels of organic contaminants via sorption into the organic carbon component. Particles that have little associated carbon, and which are largely composed of inert inorganic minerals, may not carry as much of a contaminant load. Both particle types may be present in the same soil sample and they may or may not be roughly the same size. If particle types are of the same size, sampling to produce an analytical sub-sample that is representative of the bulk average might not be difficult. Common mixing and sub-sampling techniques do not discriminate against particles if they are all the same size.

Common sample handling techniques, however, actually do separate larger and smaller particles. As sample containers are shaken, smaller particles tend to settle to the bottom and larger particles migrate to the top. This problem of segregation by particle size is exacerbated by sub-sampling utensils, such as spatulas and scoops, which discriminate according to particle size based on the design of the tool. Larger, more bowl-shaped scoops will retain larger particles that would roll off a smaller or flatter scooping surface. Sample handling and sub-sampling procedures are seldom standardized or controlled to avoid introducing particle size biases into analytical sub-samples. Laboratory technicians are likely to handle samples differently, even in accordance with laboratory protocols. As a result, obtaining an analytical sub-sample that is truly representative of the bulk average in the sample container can be a challenge. A sub-sample is sometimes more representative of the larger particle fraction, while an intended replicate sample is more

representative of the smaller particle fraction, and analytical results can vary based on the severity of this sub-sampling discrepancy.

Although soils of uniform particle size do occasionally occur naturally, soils usually contain particles that vary in size from silt- and clay-sized particles to pebbles. Clay minerals are of particular note because their large microscopic surface area and electrostatic charge tend to preferentially bind contaminants carrying a positive charge. Contaminant loading varies with particle size, and routine sample handling and sub-sampling procedures discriminate according to particle size. Therefore, soil data can be highly variable, even when extracted from the same sample jar. Laboratory duplicates, matrix-spike duplicates (MSD), and field splits are considered to be the "same" sample, yet their results commonly differ significantly when routine splitting procedures are used.

Split samples and laboratory duplicates can have very different analytical results. Both results are correct in the sense that the analysis of both analytical sub-samples was likely correctly performed; however, the analytical sub-samples are fundamentally different samples. Both could be misleading in the sense that neither data result might be representative of the true average concentration for the sample in the container or the area in the field represented by the sample. The difference between the replicate sub-samples' results is a measure of how uniform the soil concentration is at the within-sample spatial scale (the within-sample matrix heterogeneity). A few native soil types may show near uniformity *in situ*, but often, deliberate sample preparation activities are required to generate a homogenous sample. Highly variable analytical replicate results should be a warning to decision-makers that the data generation process may not be sufficient to control artifacts caused by sample heterogeneity, and decisions based on those data have an additional degree of uncertainty.

Fortunately, strategies are available to control this source of data variability, particularly if the analytical variability in sample replicates or splits exceeds what can be tolerated in decision-making. Unfortunately, options for reducing within-sample heterogeneity involve additional labor, time, equipment, and costs. The additional resource requirements, however, should be weighed against challenges arising from increased data scrutiny; the need for additional data collection to resolve data conflicts; and the need to support difficult risk management decisions concerning protectiveness, which can affect remedial designs and costs.

The exact procedures selected to prepare samples are dependent on the myriad of project variables related, but not limited to:

- soil type,
- contaminants of interest,
- staffing,
- budget,
- availability of equipment,
- desired workflow,
- number of samples, and
- subsequent sample preservation or preparation steps.

Under some circumstances, some or all of the sample preparation might be performed onsite, either in a temporarily housed fixed-base laboratory, or in a mobile laboratory. Alternatively, some or all of the sample preparation might be performed in an offsite, fixed-based laboratory. This decision should be made by the project team after deliberation of relevant factors.

There is yet another critical factor to consider when determining the actual soil sample population of interest. Sometimes the population of interest is the bulk material, meaning that the data are supposed to represent the average concentration for all the soil material (usually defined as particles less than 2 mm) in the sample. Frequently, however, this is not the case, such as when the population of interest is defined by what decisions the data are intended to support. Risk decisions may involve exposure pathways that are governed by particle size. For example, small particle sizes that may be transported as dust may be most likely to be carried off-site, also ingested, or inhaled. If the dust-sized particle fraction is the population of interest, it is inappropriate to use data generated from bulk samples. The dust fraction must be isolated from the bulk sample and analyzed separately if the sample is to be representative of the population relating to the decision process. This is an example of where a well-prepared CSM can be effective in supporting the design of the project.

Procedures that can reduce within-sample variability include:

- breaking up aggregates by hand, by pounding or by grinding in a mortar and pestle;
- sieving to a uniform particle size;

- milling using mechanized grinding equipment (not anticipated by the UFP-QAPP template); and

- using incremental sub-sampling (a "slab cake" or "pancake") to create the sample mass for analysis

Sieving requires a dry sample, so wet samples need to be air- or oven-dried. If oven-dried, the oven temperature should not exceed that which would drive off the more volatile analytes. Any analyte that is volatile enough to be run through a gas chromatograph, like the SW-846 8270 semi-volatile organic compound (SVOC) list, can be lost from a sample baked at too high of a temperature. Air-drying, therefore, while more time-intensive, limits the potential for analyte loss.

Sieving may be performed by hand or by a mechanical shaker. If done by hand, shaking should continue until separation is complete to ensure the entire fraction of interest is obtained.

Just as incremental sampling increases the representativeness of a *field* sample, incremental *sub*-sampling (using the "slab cake" technique, see attachments to UFP-QAPP template Worksheet #18, 21, 26) serves to increase the sample's analytical mass representativeness of the entire sample. Incremental sub-sampling can occur in stages. It can be performed onsite (if conditions are amenable) or in an offsite laboratory. The first stage of incremental sub-sampling

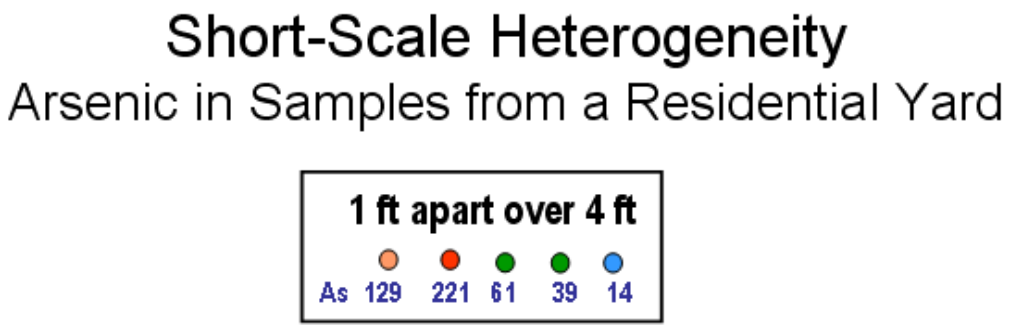
representatively reduces the volume of the original field ICS sample to a smaller volume that might be submitted to the laboratory or be used as an increment in a 2-tier ICS strategy. On-site incremental sub-sampling and laboratory sub-sampling are illustrated in the "slab-cake" photos on the left and right above, respectively.

Incremental sub-sampling suppresses data variability caused by micro-scale heterogeneity within the sample. This assumes that the population of interest has been identified and is what is being sub-sampled. It also assumes that the sub-sampling scoop has been sized to equally retain all particle sizes present in the population of interest.

Within-Sampling Unit Variability (Short-Scale Heterogeneity)

The next tier of matrix heterogeneity is within-SU variability. SU's are defined areas/volumes of soil from which increments are collected to produce a single sample representing that entire area/volume. The goal behind incremental sampling of a SU is analogous to the goal for incremental sub-sampling from a "slab-cake" created from a jarred sample, only on a larger spatial scale. Just as incremental *sub*-sampling seeks to reduce the data variability between analytical sub-samples, incremental sampling of a SU seeks to reduce the data variability between field samples intended to represent some relatively small area/volume (the SU) that is a portion of a larger DU.

Within-SU variability is a function of the short-scale heterogeneity created by deposition and transport/transformation mechanisms. Short-scale heterogeneity often exists over distances of feet to yards; for example, the spatial scale at which collocated samples might be collected. The effect of high short-scale heterogeneity is that a different concentration may result depending on which exact location is chosen for sample collection. If one location is chosen for sampling, the result may be low, but another location only 6 or 12 inches away may give a very high result. An example of heterogeneity at this scale from actual field data for arsenic in parts per million (ppm) using X-ray fluorescence (XRF) is provided below:



On the other hand, typical long-scale heterogeneity operates on a scale of yards to acres and is the heterogeneity that occurs from one area to another within the DU. The differences in contaminant concentration could be due to point sources or topography-driven transport mechanisms. Unless the average over a large area is the only measurement desired, SUs can be structured so that they avoid mixing potentially different populations represented by long-scale heterogeneity. If the DU result were to exceed the action level, and further work is required, the archived SU samples can

provide the information needed to update the CSM, design remediation, or perhaps split the DU into two or more DUs.

The benefit of SUs to risk assessment is explained in the *Risk Assessment Guidance for Superfund* (1989c):

> "If samples are taken from an area that is anticipated to have a high degree of variability in chemical concentrations, then many samples may be required to achieve a specific level of certainty and power. If contaminant concentrations in an area are highly variable and only a few samples can be obtained, then the risk assessor should anticipate (1) a great deal of uncertainty in estimating mean concentrations at the site, (2) difficulty in defining the distribution of the data (e.g., normal), and (3) upper confidence limits much higher than the mean. Identification of multiple areas of concern—each with its own set of samples and descriptive statistics—can help reduce the total variability if areas of concern are defined so that they are very different in their contaminant concentration profiles. Risk assessors should discuss during systematic planning both the anticipated variability in the data and the desired power and certainty of the statistics that may be estimated from the data."

Project teams are encouraged to explore the SU concept with their risk assessors to determine potential applicability on a site-specific basis.

SUs are also helpful when a DU is so large that coverage of the entire DU with increments is impractical. A single tier of increments over a large DU (e.g., one incremental sampling project in an agricultural setting had DUs of 80 acre size) should have an increment density that can control for both short-scale and long-scale heterogeneity at the same time. For practical reasons, increments generally have a small sample support (see definitions list #7). Sometimes the sample support may be only the dimensions of a corer 1 inch or so in diameter and several inches in length. Very small increment dimensions mean that more increments are needed to adequately capture the concentration extremes caused by a small sample mass in the presence of significant micro- and short-scale heterogeneity. SUs can be used to create larger sample supports that avoid the problems created by short-scale heterogeneity. As long as contamination is randomly distributed across the DU, the variability between SUs, that is, the long-scale heterogeneity, will be low. The number of SUs needs to be sufficient to capture and measure long-scale heterogeneity. If the variability between SUs is low, then fewer SUs are needed to represent a large DU. Statistically, the fraction of SUs needing to be sampled in a very large DU (tens to hundreds of acres) depends on how variable the concentrations are between the SUs within that DU.

In summary, as short-scale heterogeneity increases, more increments are needed within the SU. As long-scale heterogeneity within a large DU (which has been partitioned into many SUs) increases, more SUs within the DU need to be sampled. For small DUs, it is usually practical and advisable to sample all the SUs.

Using SUs has the additional advantage of being amenable to simultaneously gather information about the spatial distribution of contaminants within a DU. SU data can preserve information on spatial contaminant distribution if that information is important. Such information might be needed to affirm, refute or correct components of CSM. For example, if one SU out of four has a

concentration that is inconsistent (significantly different) than the other results, and the CSM instead assumes contamination is completely randomly distributed due to aerial deposition, the CSM is either incomplete or incorrect. Building an accurate CSM generally requires a weight of evidence approach, and a SU strategy can help provide that. It is important to note, however, that chemical data are only one line of evidence. Site visits, field observation, and other types of data such as the slope of the land, type of soil, vegetation, moisture, and wind direction provide some of the most valuable evidence. The Preliminary CSM provides the justification for designing an ICS strategy at the beginning of a project. As the project progresses, defending that the data set was collected appropriately requires that the assumptions supporting the sampling design be verified by a transparent and well-documented CSM.

Decision Units: the Core of an ICS Design

DUs are the fundamental basis for decision-making purposes under the UFP-QAPP template. DUs are designed to address EUs and exposure assumptions and pathways. All other components (SUs, etc.) fall out of the DU design. A DU is a large volume of soil, generally at the scale of acres or fractions of acres. A DU is often discussed in terms of two-dimensional area, but its depth component gives the DU three dimensions. Setting the depth is a critical part of developing the sampling design and defining the DU. As implied, a DU is that fixed volume of soil for which a primary decision is made. As an example of a primary decision: "Is the TEQ or TEQ UCL for this ¼-acre residential yard at a depth of 0 – 2 inches, above the regulatory threshold?"

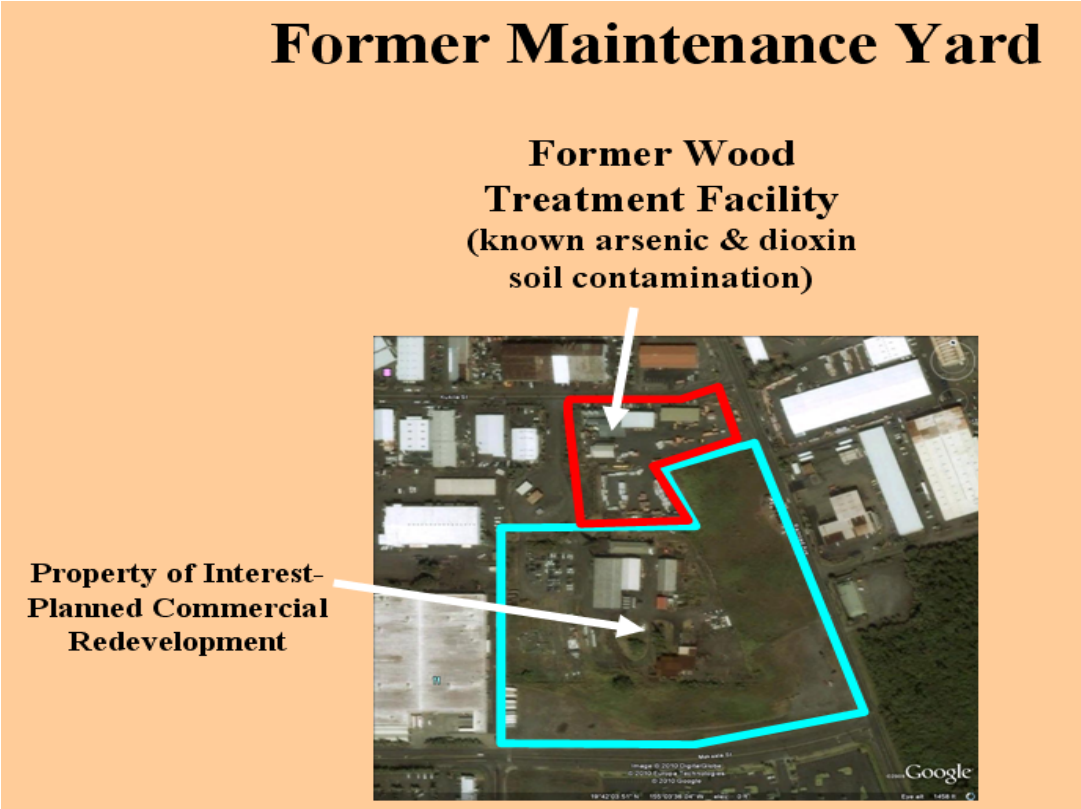
Regional technical teams have the flexibility to develop the size, shape, and orientation of DUs necessary for risk evaluations. During systematic planning efforts and in consultation with key technical team members such as risk assessors and project engineers, the team should use all available information (captured in the CSM) to define DU boundaries. Former and future land use, historical sampling data, site physical features, and other components of the CSM help define DU size, shape, and orientation necessary to complete project-specific UFP-QAPP elements.

Two brief examples are provided to illustrate how teams might use CSM information and exposure scenarios to develop appropriate DUs for ICS efforts.

DU Development - Example 1

Example 1 illustrates a commercial redevelopment area adjacent to a former wood treatment facility with known arsenic and dioxin soil contamination. DU Example 1, Figure 1a below, illustrates the shape of the area of interest outlined in blue along with the boundaries of the former wood treatment facility (outlined in red) neighboring this property.

DU Example 1; Figure 1a

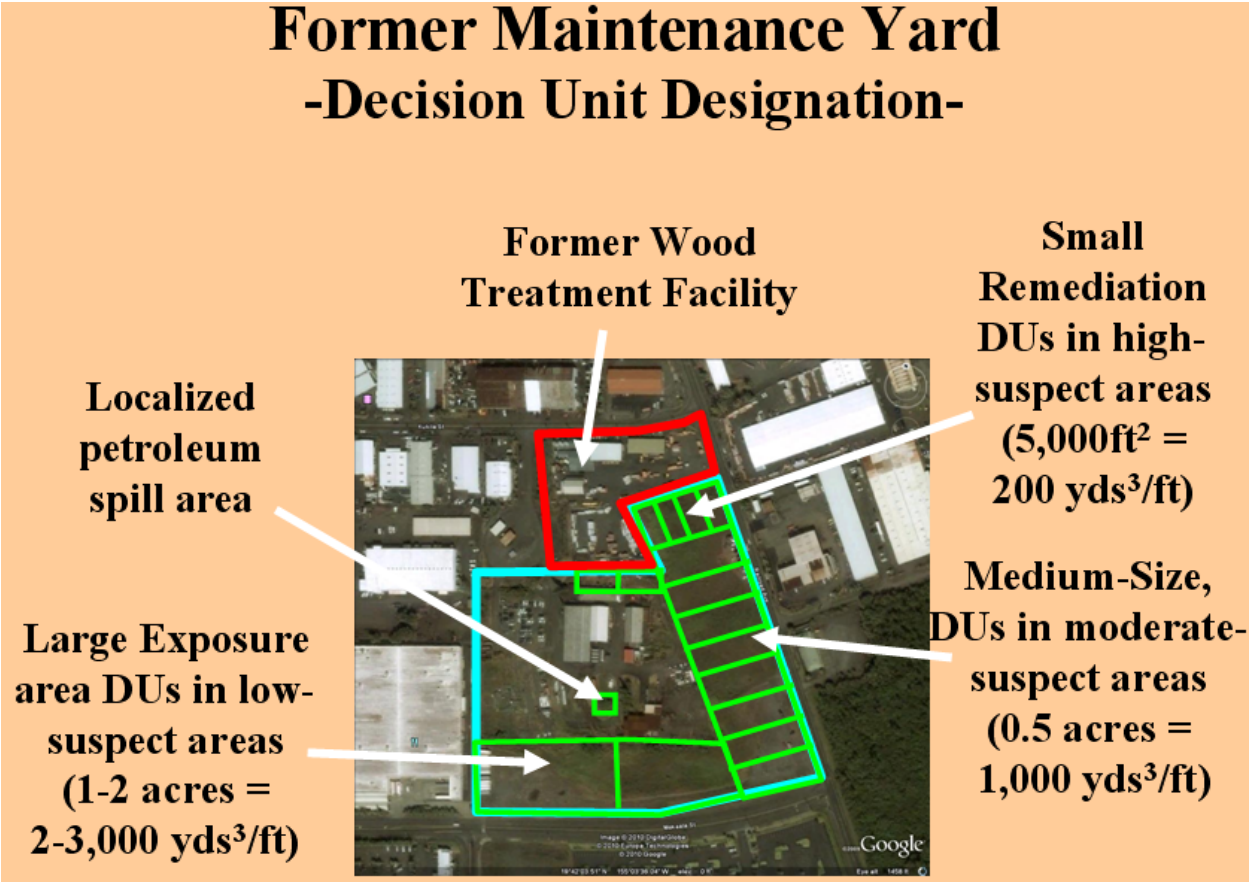


[Figures provided by Roger Brewer, Hawaii Department of Health. These examples include site work conducted EPA Region 9. Case studies illustrating these concepts were provided in a presentation to the EPA Dioxin Assessment UFP-QAPP Development Work Group on July 28, 2010.]

Illustrated below (see Figure 1b) is the same property layout, but with the DU borders drawn in. The figure's historical information and the current CSM is used to define acceptable DU size, shape, and orientation. With the DU lines overlaid, the above figure is transformed into the property shown below (DU Example 1b).

Figure 1b places smaller DUs with higher likelihood for potential remediation along borders with the former wood treatment facility and medium size units are placed along historical access points to that property. A specialized small DU is used to encompass an area of a known previous petroleum spill, and larger DUs 1-2 acres in size with a low likelihood of contamination are placed further away from the known source area.

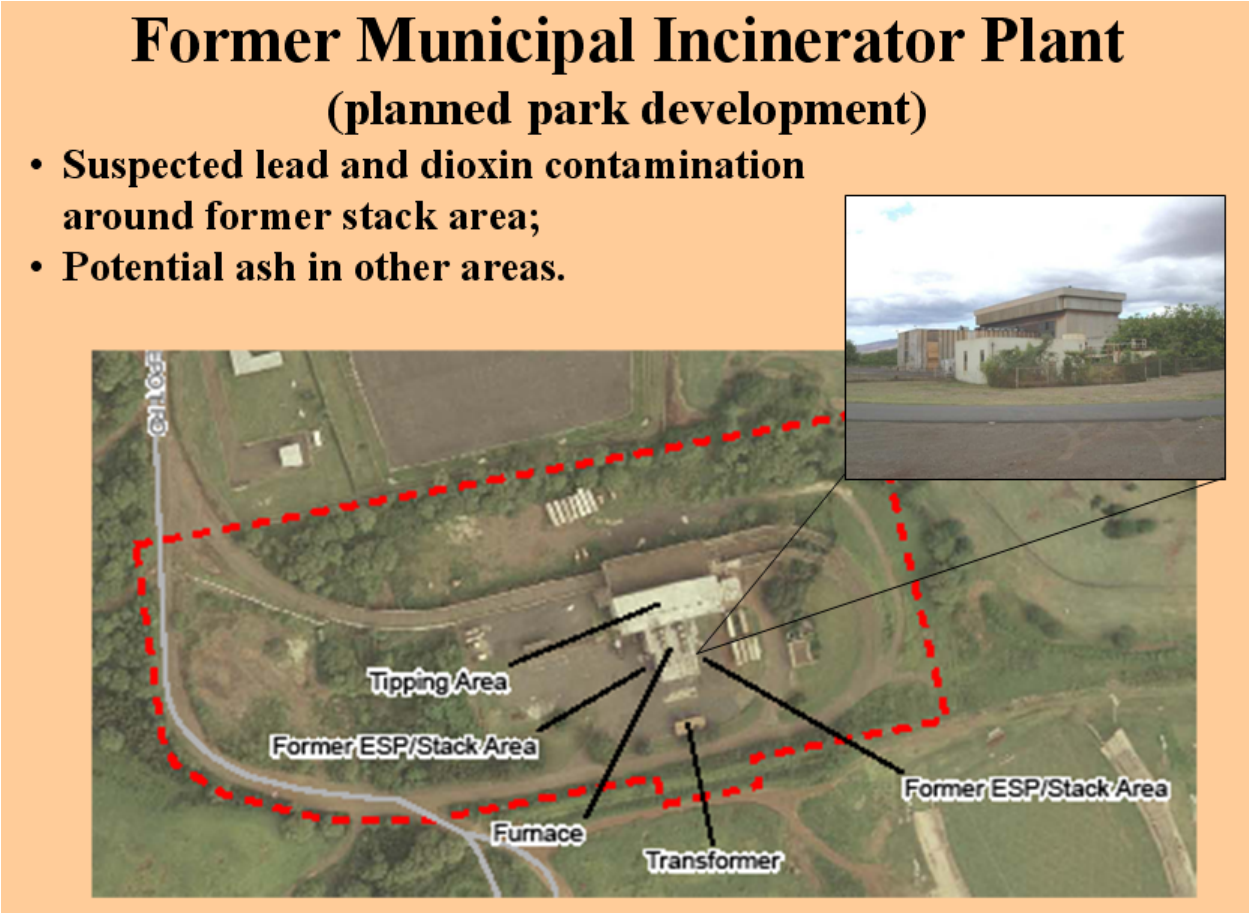
DU Example 1; Figure 1b



DU EXAMPLE 2

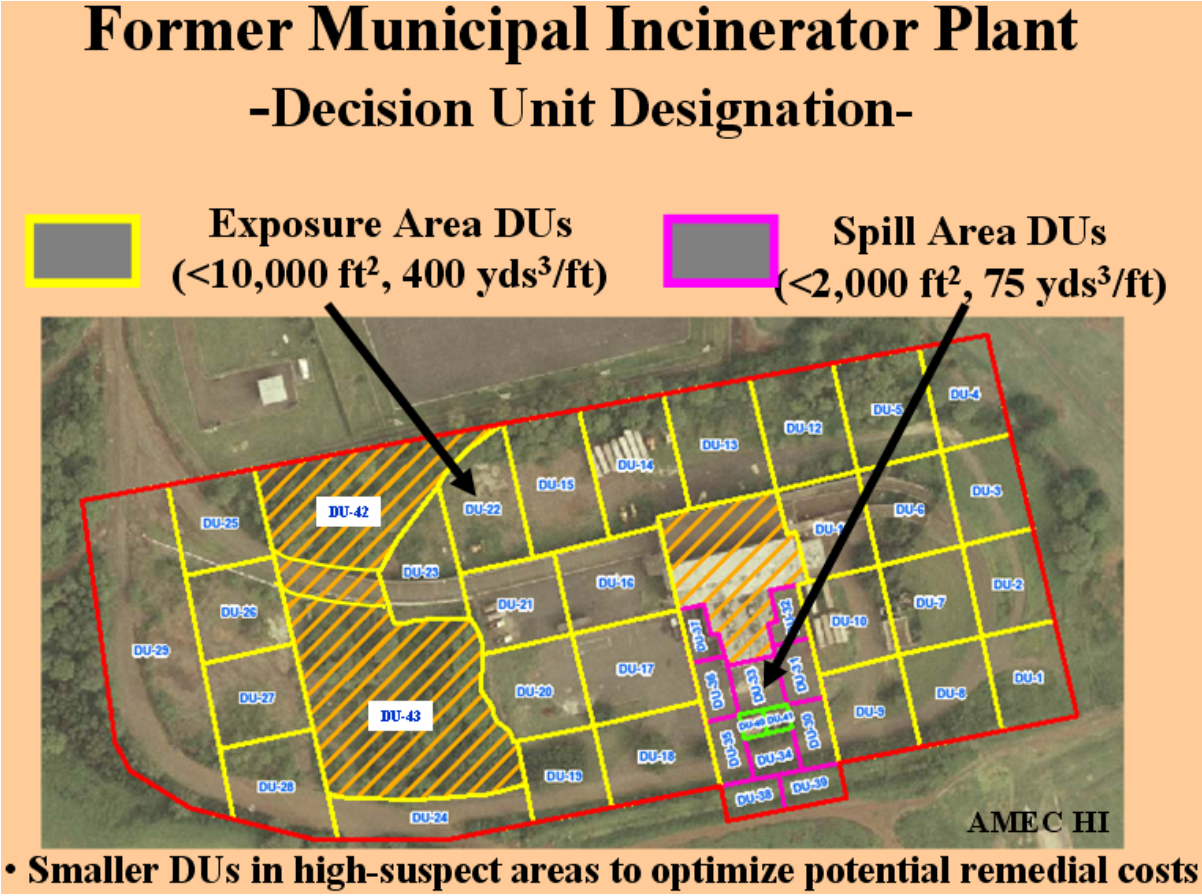
DU Example 2 below illustrates the boundaries of a former municipal incinerator with suspected lead and dioxin contamination slated for redevelopment as a park. Once again, historical information and CSM elements define appropriate DU placement, as well as their sizes, shapes, and orientations as designed during systematic planning. Areas with unique site features suggest locations with increased likelihood of contamination, for example: the tipping area, former stack, furnace, and transformer areas.

DU Example 2; Figure 2a



The figure below, DU Example 2; Figure 2b illustrates how smaller DUs encompassing potential spill areas or "hotspots" are placed around areas with the important CSM features noted in the figure above. The DUs are placed around source areas in such a way as to capture runoff impacts as well as bound the high concentration area. The DUs may be as large as needed to encircle probable contaminant populations. Those DUs are designated as "removal DUs" or "spill DUs" since their primary use is to gather information to complete remedial designs. Therefore, the approximate concentration in the soil and the extent of contamination needs to be known. A simple average concentration using a single set of 30 increments to generate a close estimate of the mean is collected for disposal pricing. Note that because these are the "spill DUs" where ICS analytical results are expected to exceed threshold criteria and not an "exposure DU", an approximate value for the mean is sufficient; as such, replicate DU-ICS samples and a UCL are not necessary to facilitate project decision making.

DU Example 2; Figure 2b



Setting Removal Unit Boundaries with Short-Scale Composites

When there is no information "on the ground" that points to where to draw the edge of a contaminant boundary, taking short-scale composites, rather than discrete grab samples, to bound excavation areas can guard against errors caused by short-scale heterogeneity and misleading data. Short-scale heterogeneity can cause a single sample to have much lower or higher results compared to its neighbor's only inches away. A short-scale composite, such as a 5-point composite over several square feet, serves the purpose of a typical grab sample, but is much less prone to the extreme variability that is sometimes seen with discrete samples. In this case, a boundary might be falsely flagged as "clean" or "dirty" based on the non-representative result. On the other hand, a short-scale composite sample can cover a 1- or 2-sq.ft. area with approximately five increments. This increases confidence in decisions made on the data point from that location. It is more likely that the single grab sample will be non-representative of that small area; whereas a multiple increment composite has a much higher likelihood of accurately representing the concentration used to decide "clean" or "dirty" at that spot.

Implementation Under the UFP-QAPP for Dioxin Soils Assessment

The definition and concepts discussed in this User Guide provide information that allows sufficient flexibility for project-specific UFP-QAPP application in a variety of Regional and Site-specific scenarios. Flexibility includes, but is not limited to: defining the size, shape, orientation and depth of appropriate DUs, defining exposure assumptions, number of increments, particle size of interest, archiving, and quality control, etc.

To illustrate these concepts in the context of the dioxin soils assessments and assist with project-specific UFP-QAPP development and application, three examples are provided, as follows:

- Example 1 illustrates the application of a simplified DU approach utilizing the default 30-increment samples. In this example, the only application of archiving is the field or analytical replicates desired by the planning team.

- Example 2 illustrates the same ICS DU approach but provides additional information on defining specialty DUs based on intimate site knowledge and a robust CSM.

- Example 3 is the most complex example drawn from work at a hexavalent chromium site. This example illustrates how DU and SU archiving can be combined to manage heterogeneity at various scales, while addressing larger DUs.

Regional technical teams have the ability to augment or adjust these illustrative designs to match project-specific needs. However, modifications should remain in accordance with the concepts presented in this User Guide.

Example 1 - Simplified Decision Unit with Default Increments with Minimal Sample Archiving

This investigation area consists of one 0.25-acre site. The DU-ICS sample is formed directly from 30 to 60 increments, there are no SUs and there is only one tier of ICS analysis. Field replicates for DU-ICS can be collected for the DU as determined during systematic planning (the procedure for collecting replicates should be described in an attachment to project-specific UFP-QAPP in Worksheet #17). An example of how technical teams may operationalize these procedures is provided below, however, steps illustrated in this example should not be construed as requirements.

At the DU:

1. Collect 30 to 60 increments for DU-ICS sample formation. For this site, because the investigation area is relatively small (0.25-acre) and previous data collected does not show extreme heterogeneity, 30 increments are selected.

2. Combine and homogenize increments to form one DU-ICS sample. All material may be shipped to the laboratory where the DU-ICS sample can be sub-sampled in accordance with UFP-QAPP procedures and remaining material can be sub-sampled for laboratory

replicates and archived. The User Guide section on "within sample matrix variability" provides a discussion of laboratory sub-sampling procedures. Alternatively, teams can sub-sample DU-ICS samples in the field and archive remaining material at a suitable facility to maintain storage and custody requirements. Frozen sample material for dioxin/furan analysis can be archived up to 1 year.

3. With only one DU at this site, the DU is designated for replicate collection. Collect the same number of increments as used in the original DU-ICS sample (step 1) via a new systematic DU grid for each of two DU-ICS field replicates. The frequency of field replicates per DU should be determined during systematic planning and described in the project-specific UFP-QAPP worksheet# 17. Land use, physical attributes and other CSM factors can be used to associate adjacent DUs (if a project involves more than one DU) with field replicates to *very similar* DUs that do not have replicates. Various options are available for how to apply RSDs from DUs having replicate DU-ICS sample sets to *very similar* DUs without replicate sets. For projects containing multiple DUs with triplicate analyses, DUs with similar CSM attributes can be grouped for comparison or the DU triplicate with the highest variance can be used as a conservative estimate. The variance determined from DUs with replicate values can be used to estimate a UCL for other DUs within the grouping. Standard deviation values from similar DU-ICS replicates can be used to calculate a UCL for a TEQ value of a single nearby DU-ICS sample. Recall that standard deviation (SD) and relative standard deviation (RSD = SD ÷ mean) are interconvertable.

4. Combine and homogenize the increments collected in step 3 to form each of the two DU-ICS field replicates. As with the original DU-ICS sample, technical teams may choose to process and sub-sample in the field, in the laboratory, or a combination of both that meets project objectives. Archive remaining material as desired, in case there is a problem with rejected congener data or sample homogeneity that requires reanalysis or further sample preparation.

5. Analyze DU-ICS samples (primary and field replicates).

6. Use analytical results to calculate a dioxin TEQ (as outlined in UFP-QAPP template Worksheet #17, Section 17.2.3).

7. Use TEQ values from DU-ICS samples (primary and two field replicates) to derive a RSD and 95% UCL. Estimates of variance (RSD) from other grouped, *very similar* DU replicates can be used to assess DU results where replicates were not collected.

8. Evaluate analytical TEQ results for the field replicate sets. Field replicates results can be used as a QC check to evaluate acceptable performance of the sampling and analysis chain. This chain includes having an appropriate number of increments and adequate homogenization in sample preparation. If the replicates do not agree within acceptable limits, the laboratory replicates (archived from the primary DU-ICS sample) can be analyzed to evaluate representative and reproducible sub-sampling and acceptable analytical and clerical performance. If the replicates (field or laboratory or both) do not

agree within acceptable limits, the problem can only be further evaluated through the variability sources QC procedure the procedure illustrated in Figure 4. While it is recognized that extensive QC and replicate procedures are likely beyond the scope of a site with a single DU, additional discussion of how replicate results can be used is provided below. These procedures may be incorporated into site activities for assessments with multiple DUs:

a. For the DU where field replicates are collected and analyzed, TEQ values can be used to calculate a UCL (calculate the RSD for use in the UCL calculation from a set of three results). Unacceptably high data variability (i.e., a high RSD) may suggest that the DU's matrix heterogeneity requires denser incremental sampling coverage to ensure an accurate representation of the DU's average. However, the source of high variability should be evaluated with a series of field and laboratory replicates (the Variability Sources QC Procedure, Figure 4). This procedure evaluates which steps in the sampling and analytical procedures are contributing most to overall variability and becomes increasingly valuable as the number of DUs and the scope of the assessment increases. If the source of variability is in sample preparation (which can be revealed through the analysis of the laboratory replicates), increasing the number of increments will not address the problem.

b. How much data variability can be tolerated in the DU-ICS replicates depends on how close the concentrations are to the action level. More data variability can be tolerated if the data results are significantly higher or lower than the action level. Statistically confident decisions can be made if the uncertainty interval is wide, but is far from the action level. If the DU-ICS TEQ result is near the action level, the uncertainty interval around it may need to be narrower so that it does not overlap the action level. Overlapping of the uncertainty interval with the action level indicates that the variability in the data set is too great to support statistically based decision-making at the selected level of statistical confidence.

c. At DUs without replicate analyses (i.e. Sites where 10% or more DUs are targeted for replicate collection based on systematic planning and field conditions), a UCL can be calculated using the area-wide variability. That is, a standard deviation (SD) can be calculated when there are at least three similar adjacent DUs, each having a DU-ICS analytical result. There are strict conditions to be met for this procedure to be used. It is mandatory for all the grouped, adjacent DU-ICS samples to be very similar in TEQ results, in their CSMs and in their physical attributes. The SD calculated on this area-wide set of TEQ DU-ICS analytical results can be used to calculate a UCL for those DUs not having a set of DU-ICS replicates.

d. Another option for accommodating statistical uncertainty, when making decisions based on a single DU-ICS analytical result, is to build the uncertainty interval <u>around the action level rather than the TEQ value</u>. This is a mirror image of the confidence interval procedures that produce a UCL. Instead of building the uncertainty interval around a DU-ICS TEQ average using three replicate DU-ICS samples, data from nearly identical DUs are used to build the statistical uncertainty

interval around the action level. Then single DU-ICS sample analytical values can be compared to the confidence interval around the action level. Those values falling above or below this confidence interval represent statistically based decision making, while TEQ results falling within the interval may require additional evaluation or be subject to Regional risk management decisions.

9. If replicate data are outside the acceptable criteria as defined in the project-specific UFP-QAPP Worksheet #17 and are not similar to RSD values seen in other available replicate sets, use the Variability Sources QC Procedure (Figure 4) to determine where the problem lies so that corrective action can be taken.

10. If analytical results are within the acceptable criteria, compare TEQ values for DU-ICS samples and calculated UCLs to decision criteria. If there is overlap of the UCL with the action level, Regional management has the discretion to decide whether to 1) try to resolve the uncertainty with additional sampling, or 2) accept the result and make a decision in the face of the statistical uncertainty.

11. Consider secondary qualitative evaluation for non-risk decisions:

    a. Use analytical results to inform non-risk related decisions of concern; e.g., remediation planning.

Example 2 - Specialty Decision Units with Default Increments Optimized Based on DMA

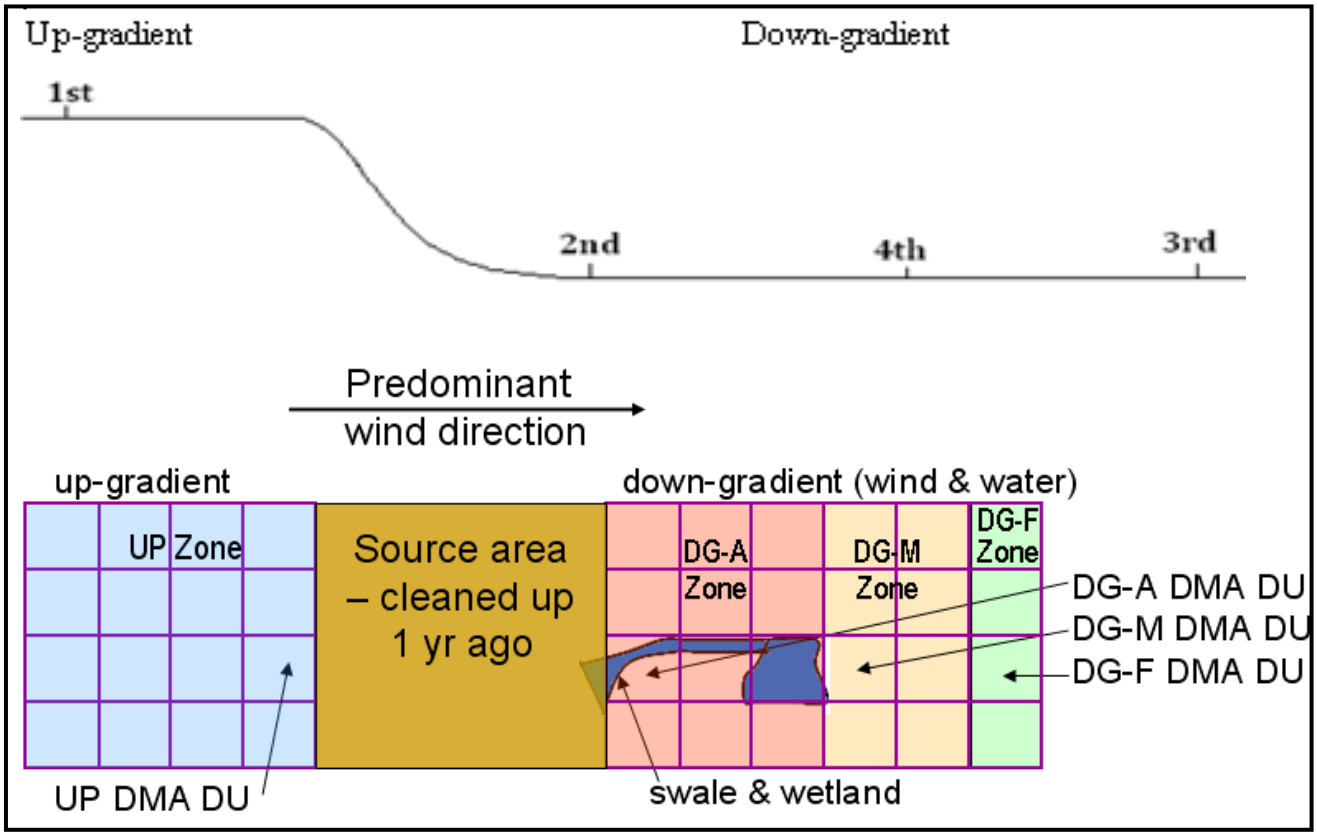
The project consists of 10 acres of undeveloped land around a 4-acre area that had been sprayed with dioxin-contaminated oil. The 4-acre source was tested and cleared after remediation was completed 1 year prior. Limited testing of the other land indicated that contamination had migrated from the source area, but the extent of migration is unknown and no remediation was performed on the remaining 10 acres. Now there is interest in developing the 14 acres for 0.25-acre residential lots. The remaining 10 acres are to be investigated prior to any development. Thus, there are 40 potential DUs to be investigated.

Approximately 4 acres of the land is uphill from the potential source area. Prevailing winds and stormwater run down a gentle slope toward the remaining 10 acres. The particle size of interest consists of any particles that could contain the contamination and be carried by wind and water onto other properties. The project team considers the CSM and historical information on prevailing winds (i.e. windrose charts), and decides to sample the top 3.0 inches of soil. The project team enlists the help of EPA's Office of Research and Development (ORD) to model contaminated particle transport under the site conditions. ORD identifies the soil fraction of interest and recommends a sample preparation technique to isolate the target soil population. A background concentration study <u>on that particle fraction</u> needs to be performed on local, but unaffected land.

The project team determines the analytical laboratory they intend to use and plan a DMA to test their CSM and the logistics and performance of the following:

- analytical procedure,
- sample collection,
- sample preparation and sub-sampling in the field and laboratory,
- sample shipment and storage,
- sample analysis, general laboratory analytical performance (from laboratory QC checks), and
- reporting of result and QC.

Site reconnaissance identifies four DUs to be sampled for the DMA. Each DU is located in an area with varying degrees of potential contamination based on the ORD modeling. The first DU is located in the upgradient area (UP) and is least likely to be contaminated by the site. The second DU is adjacent to the original site and in the downgradient zone (DG-A) which is the most likely to be contaminated by wind and water flow. The third DU is also downgradient, but is furthest (in zone DG-F) away from the site on the downgradient side. The fourth is midway between the second and third DU located in zone (DG-M) on the downgradient side.



The first of three replicate 30-point ICS samples is collected in the upgradient DU. This DU is analyzed first to test the logistics of the sample handling procedures. It is also selected as the pilot because it is likely the least contaminated and the least likely to require repeat sampling if the initial sampling procedures needs to be modified. The dry and silty-loam nature of the soil allows the samples to be prepared in the field. A large mortar and pestle is used to break up the clods. Sieving and incremental sub-sampling procedures are evaluated in the field.

After collection of the first DU-ICS sample from the upgradient DU, the field team consults with a technical member of the planning team to approve any SOP modifications necessary to optimize sampling procedures, based on initial efforts at the upgradient DU. Changes are incorporated into an updated project-specific UFP-QAPP. Revised SOPs are used to collect the remaining two DU-ICS replicates for the upgradient DU, and then for all the rest of the DMA samples. Because the TEQ for the UP-DU is expected to be low, and the distribution of contaminants (if present) is expected to be homogeneous, the team decides not to recollect the first DU-ICS sample at that time. The team believes that the minor modifications made to the sampling procedure to optimize logistics will not cause the TEQ of the first DU-ICS sample to be different from the other two UP-DU replicates. That decision can be re-evaluated when the DMA data comes back from the laboratory.

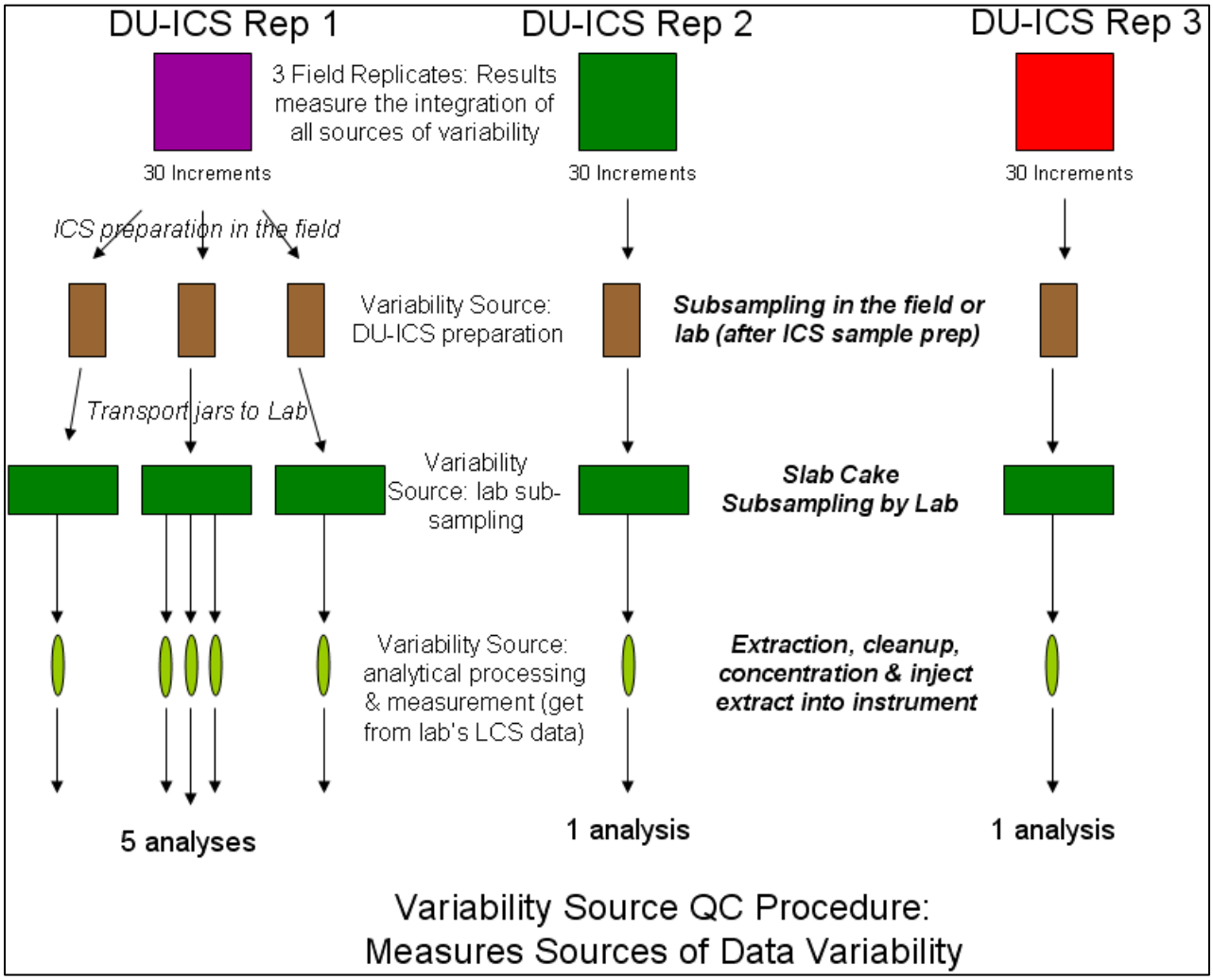
Physical observation of the downgradient DUs adjacent to the site (located in the DG-A zone) identifies a swale that collects surface runoff from the source area and drains into a 0.25-acre DU located in the DG-A zone. The swale continues through that DU, through the next DU and onto a third where it discharges into a wetland that occupies most of that 0.25-acre DU in the DG-A zone. No other DUs appear affected by swale drainage. Although the source area had been cleaned up the previous year, site knowledge indicates that the drainage channel may be contaminated enough to require remediation. The team chooses to treat the swale as a different population from the rest of the site, and the swale become its own DU. The physical dimensions of the swale become the shape of the swale DU. The swale DU itself has an area of 0.1 acres and resides within two other DUs located in the DG-A zone.

In this fictional case study, the wetland area, comprising an entire DU, will not be sampled as part of this residential reuse evaluation.

The team also desires to see what concentration levels might be detected in swale soils to prepare for the possibility of remediating this area and disposing of contaminated soil. Potential disposal costs are determined in part by TEQ concentrations found in the swale, so the team decides to collect one sample from the center line of the swale at its upper reach near the source area. This collection is part of the DMA to evaluate a worst-case disposal cost scenario. To minimize short-scale heterogeneity effects, this sample is composed of five increments taken within a 2-sq ft by 3-inch depth SU. This swale SU-ICS sample is prepared by disaggregation and sieved to less than 2 mm; the definition of bulk soil. Bulk soil is chosen as the particle size of interest because disposal of contamination from the swale would be the bulk fraction.

Two of the DUs evaluated under the DMA are at locations DG-A and DG-M. As it is likely that contamination is present in these two downgradient locations, samples are collected to complete the Sources of Variability QC Procedure (A repeat of Figure 4 is provided below for convenience). It was also possible that at least one of the two samples would have a concentration at or near the action level. It is important to characterize variability close to the action level in order to avoid decision errors when results are near the action level. For a single DU, the Sources of Variability QC Procedure requires seven analyses. Those procedures involve collecting a series of three replicate DU-ICS samples and triplicate splits.

Figure 4. Variability Source QC Procedure: Measures Sources of Data Variability (Figure Duplicated here for Reference)



Each replicate is homogenized and the particle fraction of interest isolated. The target soil fraction is placed in a large pan and a "slab cake" incremental sub-sampling technique is used to create a jar of soil to be sent to the laboratory (see the general technique depicted in the graphic below). A sampling spoon large enough to not discriminate according to particle size should used. For one of the replicates (Rep 1 in the figure below), the field slab-cake incremental sub-sampling procedure is repeated twice more to create replicate field sub-samples. The variability source QC procedure is performed for DMA DUs in areas DG-A and DG-M, but not on the farthest DU located in area DG-F since the likelihood of contamination is low. Volume for a matrix spike/matrix spike duplicate (MS/MSD) is also split from one of the replicates for analysis from the DU located in area DG-A.

All five field sub-samples in the QC set (three from Replicate 1 + one each from Replicates 2 & 3) and an MS/SD are submitted to the laboratory for analysis. In the laboratory, the five field sub-samples are again sub-sampled using the slab-cake incremental technique to get down to the analytical sub-sample mass (see figure below). One of the sub-samples from Replicate 1 is again sub-sampled in triplicate in the laboratory as part of the Variability Source QC Procedure. All laboratory sub-samples are analyzed resulting a total of seven analyses from each of the DUs in areas DG-A and DG-M.

The total numbers of analyses performed in the DMA are:

- UP DU= triplicate 30-point DU-ICS samples (three analyses)
- DG-A = seven analyses from the variability source QC procedure (+ MS/MSD)
- DG-M = seven analyses from the variability source QC procedure
- DG-F = triplicate 30-point DU-ICS samples (three analyses)
- Swale = one DU-ICS sample (of five increments) analysis
- = a total of 21 analyses plus the MS/MSD
- = total number of increments collected = 365 in 21 ICS samples

The following conclusions and CSM updates are derived from the DMA results and used to optimize the sampling design for the remaining 36 DUs plus the swale DU contained in this site:

- There did not appear to be any matrix problems for the analysis.

- The upgradient DU (UP-DU) had ICS analytical results that were detectable, but were only 1/15th of the action level value, and flagged with "J" qualifiers. The result may be within the background concentration, which can be determined by a background study that takes place during the main investigation. This result indicates how low the laboratory can report results with this soil matrix, and how precise three 30-increment DU-ICS replicates can be at low concentration, even for J-flagged values. The %RSD for the three TEQ replicates is determined to be greater than 50% (RSD= 0.5), but this is to be expected when concentrations are very low. The amount of imprecision does not require corrective action if the concentration is far below the action level because the uncertainty interval does not

overlap the action level. The within-DU variability value, however, is useful for statistical analysis, evaluating sampling and analytical performance, and formulating future strategies.

- As anticipated, the swale DU-ICS sample (which was taken near the source area-DU boundary) was elevated at 15 times the action level. The project team plans for more samples to be collected from the swale during the main investigation. Those samples are designed to evaluate whether a high concentration extends all the way to the wetlands, and finalize remedial design and disposal options for the high dioxin concentration-affected media.

- The DU-ISC samples from area DG-A have a medium-high concentration for all of its triplicate samples, at 2 to 3 times the action level. The precision for the triplicate DU-ICS samples is good enough to conclude that there is no statistical uncertainty about whether the DU is over the action level. However, the variability source QC procedure indicates that the amount of imprecision is higher than the project team is hoping to see and they note that the greatest source of data variability for the DMA sample from the DU in area DG-A is in the laboratory at the slab-cake incremental sub-sampling step.

- The DMA replicate DU-ICS samples from the DG-F DU show contamination at about 1/4th of the action level. It is possible this too is in the range of background. The results are greater than the lowest method calibration standard, and not flagged "J." The variability between replicate DU-ICS samples is similar to that seen with the upgradient DU. The 95% UCL is 0.5 times below the action level.

- The replicate DU-ICS samples from the DG-M DU have TEQ results at about 0.8 times the action level. However, with the amount of variability in the replicates, the 95% UCL for the TEQ exceeded the action level. The Variability Source QC Procedure showed that the greatest source of data variability was again in the laboratory sub-sampling step.

- Based on the conclusions from the Variability Source QC data sets, discussions with the laboratory**,** and a review of their SOPs, a request is made to the laboratory manager that a larger sub-sampling scoop (that will more consistently sample the larger particle sizes in the sieved samples) be used in all subsequent samples from this project. DU DG-M was the only DU where the decision was uncertain because the replicate mean and action level bracketed the action level. The three DU-ICS replicates were run with the new laboratory sub-sampling procedure using the larger scoop. The new mean was approximately the same, but the variability between the three results was less, although the UCL still straddled the action level. Based on the DMA result, the team defines acceptable performance of replicate DU-ICS samples as that good enough to narrow the confidence interval so that the UCL would fall under the action level when the mean of the replicates was 0.7 times the action level or lower.

- The second greatest source of data variability appears to be field heterogeneity (presumably both short- and long-scale), as measured by the three DU-ICS replicates from each of the four DMA DUs. Corrective action at the laboratory reduced variability, but if more

reduction was required, it would have to come about by increasing the number of increments per DU, i.e., have denser increment coverage.

- The concentration gradient in the "downwind" direction ranges from about 0.25 to 2.50 times the action level. EPA ORD again helped with modeling to derive zone boundaries that were 1) most strongly affected, as exemplified by DG-A (which is closest to the former source); 2) moderately affected, as exemplified by DG-M; and 3) minimally affected, as exemplified by DG-F (which is furthest from the former source). Each zone was considered separately when designing the sampling and analytical design for the main field study.

- With correction of the variability problem in the laboratory SOP, the default number of increments (30) appears to be sufficient for the areas with lower concentrations. Statistical analysis of the DMA data indicates that the 95% TEQ UCL of three replicates would be below the action level if the replicates' average value is less than 0.5 times the action level. This conclusion is derived from the variability calculated for the two low concentration DU (UP and DG-F) triplicates. This indicates that a single DU-ICS sample might be taken from these low concentrations areas, and if the ICS TEQ result is less than 0.5 times the action level, a decision of "clean" can be made with 95% statistical confidence.

- For the higher concentration DUs the situation is a little different. Revising the laboratory's sub-sampling SOP reduced data variability by about 20%. The team would like to see statistical uncertainty reduced so that when the replicate mean is less than 0.7 times the action level, the UCL also falls below the action level. To accomplish that, the team decides to control variability a little more by increasing the number of increments ONLY in those DUs suspected of being close to the action level (i.e., those in the DG-M zone).

- After considering the DMA results and conclusions, the project team decides on the following streamlined, but still conservative, sampling design:

    - For the upgradient zone (UP), which had extremely low concentrations, no replicate DU-ICS samples are planned for collection. Only a single 30-point ICS sample is planned to be collected from each of the 15 remaining DUs in the UP zone. As long as the concentration of a single DU-ICS sample is below 0.5 times the action level, a definitive decision can be made that is statistically compliant with the action level. The 95% UCL does not need to be calculated just to show compliance with the action level since the detailed work of the DMA supports that conclusion. However, if a value for the UCL is desired, the variability input (the standard deviation, SD) to calculate the UCL can be estimated using the DMA results and the between-DU variability from the UP main investigation data set.

        - There are 15 UP DUs remaining = 15 DU-ICS samples x 30 increment/DU-ICS samples = 450 increments.
        - There are 15 analyses for the UP area DUs.

- For the DG-F zone, no replicate DU-ICS samples are planned. They should all have concentrations low enough (around 0.25 times the action level) so that a single 30-point DU-ICS analytical result can be shown to be statistically less than the action level as long as it is less than 0.5 times the action level. As with the UP zone, a UCL can be calculated from the SD generated during the DMA and from the SD of the between-DUs for the DG-F zone.

  - There are three DG-F DUs remaining = three DU-ICS samples x 30 increments/DU-ICS sample = 90 increments
  - There are three analyses for the DG-F area.

- For the DG-M zone, triplicate DU-ICS samples are to be collected for every DU because 1) the concentration of DUs in this zone may be near the action level, so the team would like as much DU-specific "hard" data as possible when calculating the UCL; and 2) this is a zone where concentrations may likely show a gradient. Therefore, the concentration variability from one DU to another may not stay constant. In addition, to reduce variability in the replicate ICS analytical data set and narrow the confidence interval even more, five additional increments per DU-ICS sample are to be collected in the DG-M zone. Finally, the variability source QC procedure needs to be performed on one of the DUs chosen at random. The variability source QC does not add more increments to the collection design, but adds four more analyses. The count for the DG-M zone is as follows:

  - There are seven DG-M DUs = 7 DUs x 3 replicate DU-ICS samples/DU x 35 increment/DU-ICS samples = 735 increments in 21 ICS samples.
  - For the analysis count, there will be 21 (3 DU-ICS replicates from each of the remaining 7 DUs) + 4 analyses for the variability source QC = 25 analyses.

- For the DG-A zone DUs, the concentration is expected to exceed the action level by 2 to 3 times based on DMA results. Since the mean is already higher than the action level, it is not as important to reduce variability, and 30 increments are deemed sufficient. The wetland area within the DG-A zone is determined will not be sampled. All DUs in the DG-A zone sampled are to have triplicate DU-ICS samples. This is because the risk assessor is to perform quantitative risk assessments on the DUs in the DG-A zone. Because variability cannot be expected to stay constant from one DU to the next (because the concentration is expected to change with distance from the former source), the UCL is to be calculated from actual triplicates, and cannot be extrapolated. A MS/MSD QC pair is to be added to one of these DU-ICS samples.

  - The numbers of increments are: 10 DUs x 3 replicates x 30 increments/DU-ICS sample = 900 increments.
  - The analysis count is 30 ICS samples for 30 analyses.
  - Plus soil volume collected for a MS/MSD analysis.

- For the swale DU, the project team assumes the DU may require remediation based on the CSM and DMA data. The project team expects to collect a single swale-wide DU-

ICS sample, analyze it for all standard contaminants to prepare for disposal, and document the dioxin TEQ of the swale soil to be disposed. In addition, there is also concern that flooding may have carried high concentration material outside the visible swale boundaries. Therefore, four short scale composite samples are to be collected along the swale, two on each side. Each composite sample is to be composed of five increments within a 2-ft area to a depth of 3 inches to control for short-scale heterogeneity.

- The number of increments is 30 for one swale DU-ICS sample and 5 increments per delineation composite (2 per each side of the swale) samples = 50 increments
- There are to be five analyses (one DU-ICS sample and four short scale composites (5-increment delineation samples)

Example 2: The Background Concentration Study

Background areas are identified on all four sides around the site. As part of the main investigation, from each background area a 0.25-acre DU is selected at random to be sampled using 30-point, triplicate DU-ICS sample. The 12 ICS analytical results are to be averaged together and statistics calculated. The exact design depends on how the statistical comparison between on-site and background is structured. Guidance on the structure is provided in EPA's CERCLA background guidance (2002d).

- For background, the number of increments are 4 DUs x 3 replicate DU-ICS samples/DU x 30 increments/DU-ICS sample = 360 increments in 12 ICS samples.
- For background, the number of analyses is 12.

Adding the number of increments from the DMA to the number from the main investigation is:

- 365 in 21 ICS samples (the DMA count)
- + 450 in 15 ICS samples (the UP zone)
- + 90 in 3 ICS samples (the DG-F zone)
- + 735 in 21 ICS samples + 4 additional analyses for the Variability Source QC (the DG-M zone)
- + 900 in 30 ICS samples (the DG-A zone)
- + 50 in 5 ICS samples (the swale)
- + 360 in 12 ICS samples (background)
- = 2,950 increments in 107 ICS samples.

The total number of analyses is 111; comprised of 107 ICS samples + 4 QC samples.

Did the DMA make for extra work and cost?

If a DMA had not been performed, and triplicates had automatically been analyzed on all DU-IC samples, the number of increments would have been 40 (site) 0.25-acre DUs minus the wetland + swale = 40 + 4 (background) = 44 DUs.

- (44 DUs x 3 replicate DU-ICS samples/DU x 30 increments/DU-ICS sample)
- = 3960 + 20 (delineate the swale) = 3980 increments in 132 ICS samples + 4 five-point composites

The number of analyses = 132 + 4 (swale delineation composites) = 136

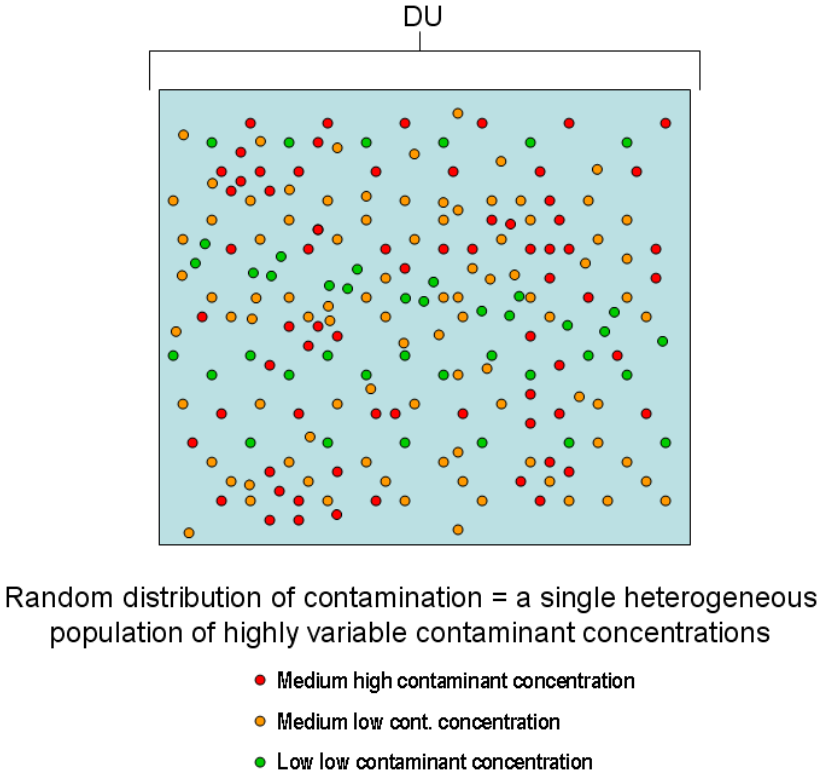
Far from causing extra work, the DMA actually saved 3980 – 2950 = 1030 increments; and saved 136 – 111 = 25 analyses. The primary benefit of the DMA, however, was the assurance that the main sampling event would generate high quality data likely to meet statistically confident decision making needs for the project.

EXAMPLE 3 - Decision Units and Sampling Units to Manage Heterogeneity at Sites with Larger Decision Units

The following is an example [based on a field project involving chromium(VI)] of long versus short-scale heterogeneity. Similar applications may be deemed appropriate for dioxin/furan projects based on Regional and project specific needs. The deposition mechanism of the Cr(VI) was manure spreaders that unevenly tossed clods of sludge onto the ground as the spreader moved along through agricultural fields. A primary transport mechanism was storm events that flushed some Cr-containing sludge and soil into lower-lying areas of the fields. Transformation of Cr(VI) to Cr(III) is influenced by a number of factors, including reduction-oxidation potential (itself influenced by moisture content, exposure to the air, and biological activity) and total organic carbon (TOC). The sludge itself held high concentrations of TOC. These transformation factors also vary from spot to spot depending on where clods landed across a sludge-treated field.

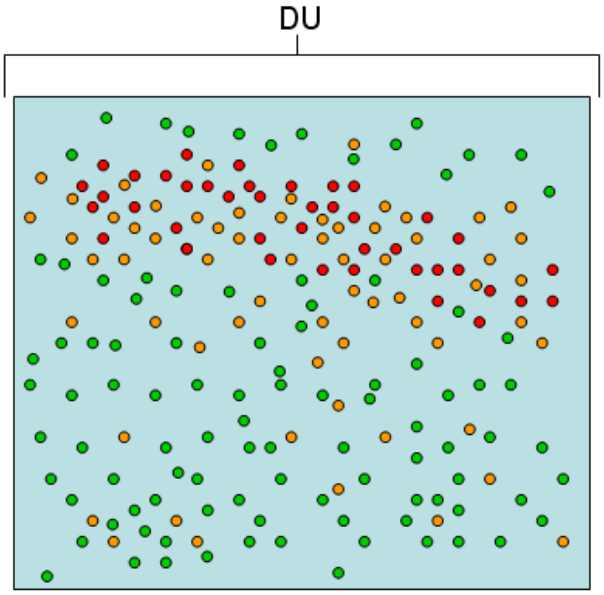
Clod deposition and upland versus lowland interactions with TOC, etc. in the soil, are suspected to have created short-scale heterogeneity within DUs.

Figure 5. Example of Random Distribution of Contamination (Figure Duplicated here for Reference)



In addition, uneven spatial distribution of Cr(VI) might be expected from location to location in response to topography facilitating transport and concentration and is an example of long-scale heterogeneity.

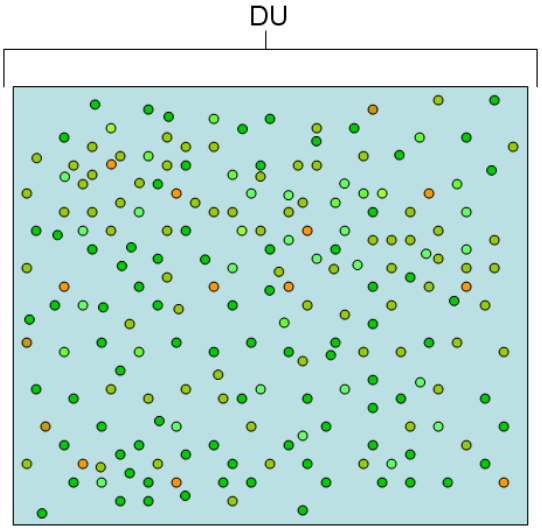
Figure 6. Example of Spatially Stratified Distribution of Contamination (Figure Duplicated here for Reference)



Spatially stratified distribution of contamination = 2 populations of contaminant concentrations as a result of a swale

On the other hand, if the Cr was *in situ* for decades in actively plowed and worked fields, these processes could have averaged out over time so that the concentration of any remaining Cr(VI) was relatively uniform or had nearly all transformed to Cr(III).

Figure 7. Example of Contaminant Distribution Overtime



Deciding between these differing CSMs required measurement of Cr(VI) concentrations from spot to spot by taking discrete samples across an area. To complete the CSM, the project team needed

to know: Were significant concentrations of Cr(VI) present in predicted locations? Was heterogeneity high or low from spot to spot (short-scale) and from location to location (long-scale)? Or was the concentration of Cr(VI) very low (having completed transformation to Cr(III) so that near-homogeneity could be expected? Higher data variability requires higher numbers of increments per SU to ensure that short-scale variability does not interfere with statistical equations. It is evident that short-scale heterogeneity is suppressed if replicate sampling over a single SU all yield similar ICS analytical results. The degree of short-scale heterogeneity was evaluated in a pilot study that at the same time tested the performance of the proposed analytical technology.

The SU size (1 acre) for the Cr(VI) project was selected based on both convenience and features of the CSM. The DU size was 80 acres based on the exposure scenario for the farmers working these fields. Judgmental selection of SUs was needed in the DMA to facilitate evaluation of the competing CSMs and to test other assumptions supporting the sampling design. Although some SUs in the main investigation could be chosen at random, the need to evolve the CSM required non-random selection. Through a “best case/worst case” strategy, a statistically conservative approach was used, as outlined below:

- Selection of a "best case" SU [having conditions favoring low Cr(VI) soil concentrations] and a "worst case" SU [having conditions favoring higher Cr(VI) concentrations] was done judgmentally.
- Putting the two data sets [one with low and one with high Cr(VI) concentrations] together for statistical analysis would create a "worst case" variability (high SD) to use in statistical equations to determine the number of increments.
- The magnitude of the variability within the SU influences how many SU increments should be taken to ensure the SU-ICS sample is representative of the SU’s true average. Therefore, this “worst case/best case” procedure, which "errs on the side of caution," was designed to guard against too few increments being collected for any SU within the 80-acre DU.
- Analytical testing dollars were saved in the pilot study because only two SUs need to be sampled, rather than the many needed for a purely statistical design.
- The pilot study also contained the opportunity to test whether the routine laboratory service provider would be able to meet the detection capability this project needed. The results of the pilot showed that it did not. Therefore, a laboratory that could perform well on selected archived soils was chosen for the main field effort.
- Many features of the pilot study were useful, and were retained during the main field effort, as follows:

    - Because of the relationship between total Cr concentration (measured by X-Ray fluorescence (XRF)) and Cr(VI) concentration by laboratory analysis, the field portable XRF (FPXRF) analytical instrument was used to screen each SU before sampling. The screening procedure measured the variability of total Cr, and that variability was calculated into the postulated variability of Cr(VI).
    - The postulated variability of Cr(VI) was used in conjunction with a spreadsheet data analyzer and VSP to determine the optimal number of increments to collect for the Cr(VI) SUs’ ICS samples.

- Similar calculations were performed to determine statistically what fraction of SUs (out of the 80 SUs per DU) needed to be sampled per DU. Because of the low concentrations and variability of Cr(VI), this ran between three and five SUs per DU. To ensure conservatism, SUs that appeared to have soil conditions favoring the survival of Cr(VI) were targeted.

Direct measurement of within-SU variability can be accomplished by analyzing individual samples from across the SU, and calculating the statistical variability for that particular sample support. However, the number of costly grab sample analyses (7 to 10 or more) to estimate variability in the SU population can be reduced if, instead of measuring the within-SU variability directly (via discrete samples), the STRENGTH of its INFLUENCE on an ICS design is measured. This is done by replicate ICS sampling within the same SU (this same principle also applies to DU-ICS replicates) to see how similar the results are. In other words, if an adequate number of increments is chosen to sample the SU, and the other sampling and analytical variables have been controlled, the RSD between the three replicate results is expected to be low. If the RSD is not low, this is an indication that 1) there are too few increments; 2) the variability in concentration within the SU was large; and 3) three replicates might be too few to capture and identify the source of variability to update the CSM.

Based on DMA results from the two SUs sampled and subsequent data analysis from the full-scale application, the technical team made the following observations.

1. The new laboratory performed the Cr(VI) analysis well, meeting analytical objectives.
2. As it turned out for the project, the overall concentration of Cr(VI) was extremely low, and far below the Cr(VI) action level. The study also found that Cr(VI) concentrations were also fairly consistent from one soil sample to another, from one SU to another SU, and from one DU to another DU. Therefore,
    - the number of increments needed to adequately represent each SU was low;
    - the number of SUs needed to represent each large DU (comprised of 80 SUs) was low, and
    - three replicate SU-ICS samples were sufficient to calculate confident 95% UCLs, each composed of that number of increments, collected from the SU.

3. In all fields tested, the amount of variability in Cr(VI) concentrations across the SUs and DUs was low, as well as the concentrations themselves. Chemical analysis and physical observations indicate that co-deposition with high TOC material speeds the conversion of Cr(VI) to Cr(III).

4. The pilot study had looked at the relationship between the concentration and variability between total Cr (mostly Cr(III)) and Cr(VI). It was quickly determined that there was an approximate relationship. This allowed the FPXRF to be used in the field for several purposes:

    - The FPXRF could measure total Cr and then a mathematical conversion could estimate the amount of Cr(VI) present. This information was used in real-time to determine in the field whether any particular DUs or SUs could have unexpectedly

high levels of Cr(VI). If such a field was encountered, the sampling design would be adapted by increasing the number of increments collected from a SU, and/or by increasing the number of SUs sampled within a single DU.

- The FPXRF was also used to estimate the thoroughness of sample preparation before samples were packaged for transport to the laboratory. If the FPXRF showed excessive variability in total Cr (via replicate measurements in the sample bag), it was reasoned that Cr(VI) might also be poorly homogenized. Additional homogenization, therefore, was applied to the sample.

5. Later, when Cr(VI) results came back from the laboratory, the assumptions and strategies used to design and implement the sampling design were borne out to be true, as follows:
    - All Cr(VI) concentrations were well below the action level; there was no chance for a UCL to exceed the action level.
    - In general, the degree of homogenization at the laboratory sub-sample level was better for total Cr than the same degree of homogenization for Cr(VI); this is a function of the much smaller concentrations of Cr(VI), and perhaps its association with soil micro-structure.
    - QC procedures found that sources of variability for Cr(VI) were split between micro-scale, short-scale, and long-scale. In general, the pattern showed increasing variability going from micro-scale, to short-scale, and to long-scale heterogeneity. Occasionally, micro-scale would exceed the other sources, indicating Cr(VI) no doubt can be intimately associated with finer particles, such as organic carbon.

Appendix 1: Best Practices Inventory for Implementing and Assessing Elements of a Successful Project
Appendix 2: Rationale for a Default of 30 – 60 Increments
Appendix 3: Demonstrations of Methods Applicability under a Triad Approach
Appendix 4: Calculation of Total Dioxin TEQs with Nondetect and Rejected Congeners
Appendix 5: References

APPENDIX 1

Best Practices Inventory for Implementing and Assessing Elements of a Successful Project

Cover Page

The tasks listed in this checklist describe activities at the implementation level, therefore the list is very detailed. It is hoped that the user finds this level of detail helpful, however not all activities in the checklist are required for a project to be successful. Each project is unique and technical teams are encouraged to evaluate these best practices for application on a project-specific basis. While all of these activities have been found to be beneficial in systematic planning and project implementation, project teams may prioritize these activities to focus on areas of uncertainty with a high likelihood of occurrence or potentially significant project impact should they occur.

The checklist helps to document a concise summary of project activities that contribute to the transparency of technical project quality and management of project decision uncertainty. Activities are broken out into concrete details, and cover a wide array of actions used to plan a project, and then gather and use information to support decisions made over the life of the project. The activities listed are recognized as standard practice and/or are recommended in EPA policy.

This checklist has multiple uses. Perhaps the most beneficial use of this checklist is as a memory aid during project planning, implementation and documentation to make sure nothing of substance in site decision-making is overlooked. The checklist format also can be used to record that all potentially applicable best practices have been considered and/or completed over the life of a project.

Reviewers of project-specific UFP-QAPPs can use the form to provide feedback as to whether the indicated item appears to be missing, has been addressed but in a way that is unclear, or is present and satisfactorily described.

Best Practices Inventory

Project Planning Best Practices	Absent	Present, but unclear	Clearly written
GENERAL PLANNING ITEMS GP1) Consideration of **end-goals/preferred reuse** of the site given site owner, stakeholder, and regulatory interests.			
GP2) **Collaboration** among all interested parties (including financial institutions, contracting & legal staff) throughout the planning & implementation lifecycle. Representation of **all relevant science and engineering disciplines** during planning.			
GP3) Discovering and articulating **decision uncertainties** that need to be managed to have a successful project			
GP4) Creating a **detailed Preliminary CSM** and updating it over the project life-cycle			
DETAILED PLANNING ITEMS DP1) **SOCIAL CAPITAL** (trust, open communication, cooperation, respect for other parties' interests)			
DP2) **OUTREACH** to all appropriate parties/stakeholders			
DP3) **EXPERIENCED STAFF** with the required expertise have been identified and accessed			
DP4) Clear consensus on the desired project **OUTCOME**			
DP5) **Preliminary CSM** developed from existing information and updated as more information obtained			
DP6) Clear articulation of regulatory, scientific, social and engineering **DECISIONS** supporting desired outcome(s), which include life-cycle site planning			
DP7) Articulate the **UNKNOWNS** (uncertainties) that inhibit confident decision-making			
DP8) General and (later) detailed descriptions of **STRATEGIES** to manage those decision unknowns/uncertainties			
DP9) Describe **INFORMATION** gathering/generation techniques (e.g., records reviews, interviews, groundwater flux testing, photo analysis, chemical data generation, geophysical tests, etc.) to manage unacceptable uncertainties to an acceptable level.			
DP10) Develop a **WEIGHT OF EVIDENCE** approach that explains which information is to be used to manage which decision uncertainties			

Project Planning Features (cont'd)	**Absent**	**Present, but unclear**	**Clearly written**
DP11) Acknowledgement of potential and actual sources of **INFORMATION UNCERTAINTY** and relevance to decision confidence			
DP12) Information acquiring technique selection is guided by evaluation of the **COST-BENEFIT** value of the information.			
DP13) Identification of **REGULATORY** authorities/ARARs			
DP14) Project **FUNDING** and **CONTRACTING** mechanisms; monitor budget status			
DP15) Known and potential **RPs** and legal considerations			
DP16) Include the costs of environmental **INSURANCE** and redeveloper risk (and how decision uncertainty affects both) on the site's lifecycle costs			
DP17) Assess whether the project (or parts of the project) can benefit from **REAL-TIME DYNAMIC/ADAPTIVE** work strategies or whether such strategies are feasible			
DP18) Ensure planning process is well-documented in acceptable **WORK PLAN** or UFP-**QAPP** formats			
DP19) Outline the communication and documentation process for recording and justifying when there are substantial implementation **DEVIATIONS** from that written and approved in the planning documents.			
DP20) Plan for on-going documentation of the information materials that would be included in a structured **CASE STUDY** write-up.			

CSM Features (a systematic planning activity)	**Absent**	**Present, but unclear**	**Clearly written**
CSM MATERIALS clearly describe:			
CSM1) known and suspected contaminant sources, release mechanisms, and amount released;			
CSM2) fate (including degradation products) and transport/migration mechanisms;			
CSM3) known or suspected contaminated media (waste, soil, GW, SW, sediment), spatial/temporal boundaries, and define at least two (and probably more) separate contaminant populations in the context of the project's intended decisions;			
CSM4) likely interactions between contaminants and matrix constituents;			
CSM5) degree of contaminant heterogeneity (contaminant distribution) at long-, short, and within-sample spatial scales;			

CSM Feature cont'd	Absent	Present, but unclear	Clearly written
CSM6) evaluate degree of mismatch between matrix variability, decision support & the sample support of anticipated sampling and analysis techniques;			
CSM7) known or potential reuse options, prioritize according to stakeholder wishes, expected site conditions, and the projected cost to achieve 1st choice, 2nd choice, etc.			
CSM8) known and potential exposure pathways and receptors;			
CSM9) probable remedial, redevelopment or Institutional Control options to achieve site reuse and reduce/eliminate receptor exposures.			
CSM10) determine decision support for each characterization, exposure, remedial or compliance decision			
CSM11) determine proper sample design, collection and handling techniques to tailor sample support to be representative of the various decision supports			
CSM12) consider what graphical or mapping techniques may be used to display chemical data and other information comprising the CSM in a form that is easily understood			
CSM13) continually re-evaluate what information is needed to guide selection, design and operation of effective remedial techniques			
CSM14) predict what information the CSM can provide when it is mature enough to support each decision			
CSM15) **CSM UPDATES** are recorded regularly in project documentation, and the information passed onto future teams involved with the site.			

Reducing Uncertainty through Information *Note: "Information" includes both measurement (chemical and non-chemical) and non-measurement information*	Absent	Present, but unclear	Clearly written
GENERAL INFORMATION ITEMS GI1) Work plan shows understanding that **decision uncertainty** is managed by using/gathering the best available information as efficiently and economically as feasible.			
GI2) Clear articulation of **what information is needed** to reduce the risk of making the wrong decision(s) (i.e., manage decision uncertainties).			
GI3) Match decision uncertainty management to various **management strategies**: 1) better utilize existing information; 2) gather new non-measurement information; 3) gather new measurement information. (See below)			

GENERAL INFORMATION ITEMS cont'd	**Absent**	**Present, but unclear**	**Clearly written**
GI4) Evaluate all sources of **information uncertainties** (for both measurements and non-measurements. For measurements, consider the effect of heterogeneity and the appropriate uses and limitations of statistical analyses and other descriptive and predictive models.			
GI5) Employ **collaborative data set**s (when analyzing the same analytes by different analytical chemical methods) and **weight-of-evidence** approaches (to blend different types of information into the CSM) to manage various kinds of information, uncertainties and data interpretation.			
DETAILED INFORMATION ITEMS DI1) Articulate decisions to be made to achieve **site reuse** or other long-term goals for site (not just for this project) and what general information is needed to achieve site goals.			
DI2) Use long-term goals as an anchor for defining what contribution **this project** can make for achieving long-range site goals. Determine what **general information** is needed to make project decisions that achieve project goals.			
DI3) Over the course of project planning, break general information needs down into the **specific information** needed to address specific project decisions.			
DI4) For each type of information, consider whether the project team is unsure about the reliability of that information (i.e., are there **uncertainties in the information**?).			
DI5) Evaluate whether some decision uncertainties can be managed by accessing **existing, but under-utilized, information** (i.e., non-measurement & measurement data).			
DI6) Evaluate whether some decision uncertainties can be managed by **gathering new non-measurement information**, e.g., Find out what stakeholders' interests are regarding site reuse. Are there new budget priorities? Could legal, regulatory, insurance and lender negotiations affect project goals and decisions, decision transparency and confidence level?			
DI7) Evaluate whether some decision uncertainties can be managed by designing efficient strategies to generate and interpret **new measurement information** (both chemical & non-chemical, e.g., contaminant concentrations, DO, hardness, GPS, geophysical, geotechnical, ecological, …), while avoiding duplication and non-informative data.			

Specialized Information in the Form of Chemical Concentration Data: **PLANNING FOR DATA COLLECTION**	**Absent**	**Present, but unclear**	**Clearly written**
DETAILED DATA ITEMS DD1) Consider the **RANGE OF DATA** generation and management options available and applicable for a weight-of-evidence approach that includes field analytics, *in situ* sensing systems, geophysical and geotechnical tools , traditional laboratories, locational, etc., and computer systems/GIS that assist project planning and data storage, display, mapping, statistics and sharing/transfer.			
DD2) Each kind of data to be collected should be matched to the data needs identified to support **INTENDED PROJECT DECISION(S)**			
DD3) Include all potential **DATA USERS** (such as risk assessors, statisticians, legal staff, etc,) when planning data collection			
DD4) Design analytical **INSTRUMENT** usage consistent with instrumental strengths and limitations			
DD5) Estimate the amount and kind of QC required to meet various data quality requirements on a decision-by-decision basis and be prepared to modify QC based on increased or decreased QA needs to accommodate matrix, method and decision situations **(ADAPTIVE FOCUSED QC PROTOCOLS)**			
DD6) List analytical QC checks and the **CORRECTIVE ACTIONS** to take when a QC check fails			
DD7) Continually assess whether real-time data is **CONSISTENT WITH THE DEVELOPING CSM**; if not, increased QC, data variability source and/or data density may be needed to ensure data & CSM confidence			
DD8) A demonstration of methods applicability (**DMA**) (attached to or detached from the main field work mobilization) is performed where performance of sampling and analysis tools is in doubt. Use **DMAs** to optimize field tools, their implementation, data management, and work flow. Also use DMA results to evaluate and optimize strategies for data generation, QC, information sharing and info management for their ability to **support real-time decisions.**			
DD9) **CONTINGENCIES/**back-up plans for sampling, analytical & software equipment failures			
DD10) **SOURCES OF UNCERTAINTY** in data are explicitly discussed and the partitioning between sampling vs. analytical variability/uncertainty is predicted and re-evaluated during refinement of the data collection process			

Planning for Data Collection cont'd	Absent	Present, but unclear	Clearly written
DD11) CSM features are used to help sampling design and handling and choice/combination of analytical options (constructing **COLLABORATIVE DATA SETS**)			
DD12) Mechanisms are developed to evaluate **DATA COMPARABILITY** for using collaborative data sets			
DD13) DATA QUALITY TERMS ("screening data quality" and "definitive data quality") are used consistent with the language of managing decision uncertainty			
DD14) DATA REPRESENTATIVENESS articulated in terms of "representative of <what matrix> in the context of <what decision>"			
DD15) Include data-related **EXPERTS** (laboratories, field analysts, instrument vendors, field-samplers, GPS, software users) in up-front project planning			
DD16) Consider the **THROUGHPUT** of sampling and analytical techniques when predicting field time required and on-site analysis costs			
DD17) Evaluate sampling/analysis costs on a life-cycle basis rather than just a **PER-SAMPLE COST** basis			

Real-time Dynamic/Adaptive Work Strategies	Absent	Present, but unclear	Clearly written
GENERAL ADAPTIVE ITEMS GA1) Evaluate whether a dynamic/real-time field decision strategy can improve the **quality of the project, while saving time** and **money.** Evaluate whether the needed **expertise and equipment** are available.			
GA2) Lay out the adaptive work strategy and logic in **decision trees (or similar mechanism)** that also address contingencies, and obtain stakeholder/regulator buy-in.			
GA3) Structure the work plan to **accept real-time information/input** from regulators and stakeholders in response to further refinement of the CSM and project progress.			
GA4) Use real-time strategies to efficiently **ground-truth & evaluate the performance of predictive models** before accepting model predictions.			
DETAILED ADAPTIVE ITEMS **DA1) CAN A DYNAMIC STRATEGY OFFER BENEFITS?** For this project consider whether moving some decision-making to the field improves decision confidence, speed site resolution and reuse, and reduce site life-cycle costs			

Real-time Dynamic/Adaptive Work Strategies cont'd	Absent	Present, but unclear	Clearly written
DA2) The **INTENDED OUTCOMES** and goals desired from the dynamic activities are clearly described			
DA3) The real-time decision-making strategy is **CLEARLY DESCRIBED** in flow charts, decision trees, tables or text.			
DA4) The dynamic decision strategy (e.g., the decision tree) is **APPROVED** by regulators in writing			
DA5) Must have property access to make a DWS work. Need physical access to offsite properties (i.e. access agreements) if plan to search for sources or define extent. Define the spatial boundaries of field work			
DA6) A mechanism is provided to easily access **OFF-SITE EXPERTISE** when needed			
DA7) The dynamic strategy includes descriptions for **HANDLING LOW PROBABILITY EVENTS OR "SURPRISES"** outside the scope of the approved decision trees; for example, when to stop work to consult with regulators and stakeholders about future direction			
DA8) Real-time compilation of data and incorporation of new information into the **EVOLVING CSM** during the project life-cycle			
DA9) Website or other mechanisms to facilitate **REAL-TIME DATA SHARING** and activity updates with regulators and other stakeholders so they can closely follow field progress			
DA10) Develop an adaptive strategy for data management, i.e., what is to be done if data transfer, storage or mapping tools do not function as intended. Ensuring the performance of data handling tools should be considered when designing the DMA.			
DA11) Make preliminary or unpolished information available to stakeholders wishing to see it; trust built through **TRANSPARENCY**			
DA12) Accommodate **REAL-TIME STAKEHOLDER INPUT** into field implementation to build confidence/comfort with the process or address new concerns as they arise			
DA13) Extend dynamic strategy to include **REMEDY-RELATED DECISION-MAKING** to degree feasible			
DA14) Consider whether a dynamic strategy can test and refine 1) fate and transport, and 2) exposure pathway **MODEL** assumptions in real-time and improve the model's predictive performance			
DA14) Consider whether a dynamic strategy can improve **CLEANUP** PERFORMANCE of on-going remedial systems while reducing O&M costs			
DA15) Consider whether a dynamic strategy can **MONITOR FOR FUGITIVE EMISSIONS** from remedial systems to ensure stakeholder comfort with process safety			

APPENDIX 2

Rationale for a Default Collection of 30 to 60 Increments

1. Project performance goals (DQO "limits on decision errors") that have a false clean error rate close to 5%.

- The Soil Screening Guidance (SSG) recommended 24 increments per a 0.5-acre decision unit (DU) based on simulations that found there was a false clean (i.e., false negative) error rate of 8% (when the true concentration is two times the Preliminary Remediation Goal (PRG) when variability over the DU equals a coefficient of variation (CV) of 2.5. (USEPA, 1996a).
- As the degree of variability (CV) rises, the number of increments to stay close to the 5% rate also rises. At a CV of three, the false clean error rate was 5% for 32 increments per DU. At CV = four, 36 increments gave an 8% false clean error rate.
- Since the commonly used DU area in residential areas today is half (i.e., 0.25-acre) that of the simulations used in the SSG (0.5-acre in the 1980's), the *density* (closeness) of increments is twice that of the SSG's. Thus, even with a high variability of CV = four; 30 increments appears adequate.

2. The precedent of recent research performed on a limited number of site types.

- One of these studies is an Interstate Technology and Regulatory Council (ITRC) study involving arsenic applied as a pesticide to a golf course.
- The results of the study showed that the mean analytical values of ICS samples comprised of 30 vs. 100 increments were essentially the same.
- However, the variability (as standard deviation, SD) between replicate ICS samples was higher for the 30-increment ICS samples. This is an expected result, since the greater the density of increments, the closer to the true mean each ICS is expected to be, and the "tighter" (i.e., less variable) the ICS analytical results are expected to be around the true mean.

3. It is conventional knowledge that "if the underlying distribution is lognormal, then the composites, viewed as physical averages, are not lognormal" (USEPA, 1996b). The more increments that comprise the ICS samples, the more normal the data distribution of ICS samples is expected to be. EPA has recognized that non-normal distributions can be normalized by using compositing strategies. The question is how many increments are needed for distribution normalcy to occur. The exact answer is that every site would have its own optimal number of increments from which a normal data distribution would emerge. But extensive customization of environmental procedures to find "exactness" often is unnecessary to get the right answer to an environmental exposure or cleanup question.

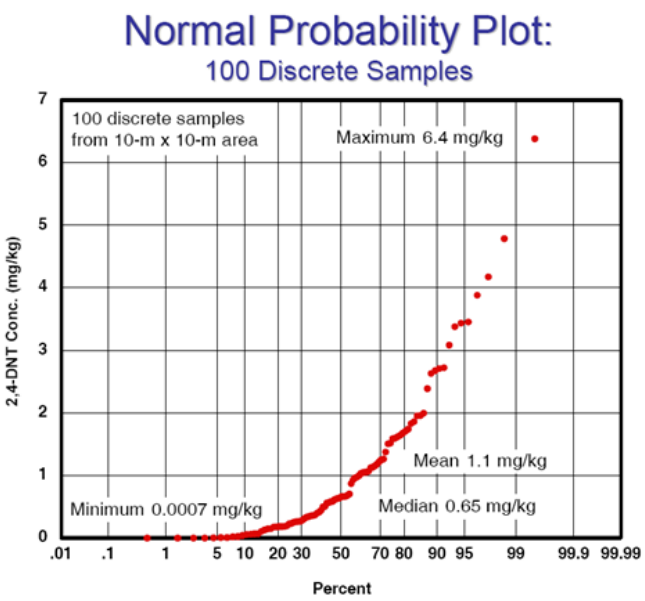
4. The hypothesis that 30 increments is often sufficient to normalize non-normal data distributions is based on the Central Limit Theorem (Arjomand, 1997). If the underlying population is only mildly skewed (non-symmetrical), fewer than 30 would suffice. If that information is available from historical data or a pilot study, then statistical calculations may predict that fewer than 30 increments is acceptable. Since this information is often not available, a default is needed. There is research data to support selection of 30 for contaminated sites. (Hewitt, et al 2009). The graphs

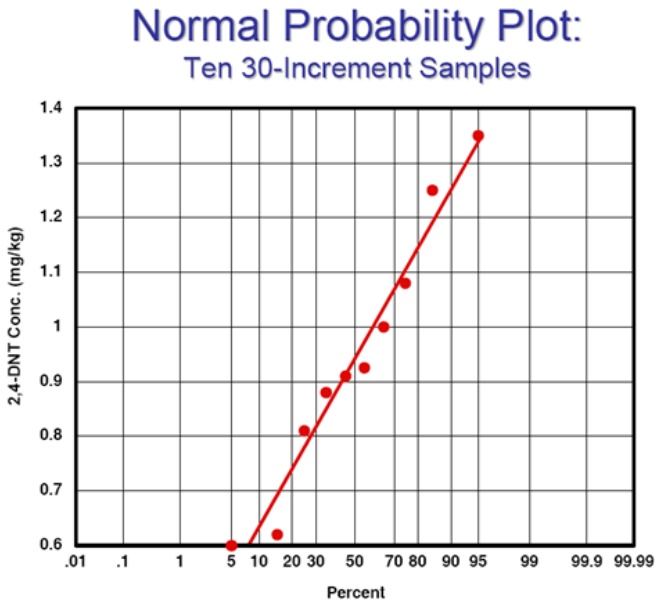
below show the results of a United States Army Corps of Engineers (USACE) study comparing discrete and incremental sampling for 2,4-dinitrotoluene (2,4-DNT), a degradation product of trinitrotoluene (TNT) that is extremely heterogeneous when residues are present in soil. As normal probability plots, these produce a straight line when the statistical distribution is normal.

The two graphs were generated using data derived from sampling the same 100 square meter area using both discrete and incremental sampling approaches. These graphs illustrate the extreme variability among the discrete analytical results, which ranged from 0.0007 to 6.4 parts per million (ppm). The distribution is clearly non-normal. The median is 0.65 ppm, which indicates that the majority of samples had concentrations below 1 ppm. The consequence is that any single discrete sample is more likely to have a result closer to 0.0007 ppm than to 6.4 ppm.

In contrast, the analytical concentrations of the 10 replicate 30-point ICS samples ranged from 0.6 to 1.35 ppm, and the distribution of those 10 results is almost perfectly normal because short-scale heterogeneity has been controlled. (Slides adapted from presentation to 2008 ITRC Spring Meeting by Alan Hewitt.) The mean values of the dense discrete data set (1.1 ppm) and the incremental data set (0.94 ppm) matched well; however, their median values did not. In a normal distribution, the median matches the mean. For the discrete 2,4-DNT data, the median was roughly half of the mean (0.65 ppm), reflecting the data's lognormal distribution. For the 30-point incremental data set, the median was 0.92 ppm, almost exactly the value of the mean. Closeness between the median and the mean is a useful indication of a normally distributed data set.

Using the Central Limit Theorem and the assumption that a set of at least 30 increments produces a normal or near-normal concentration distribution, this allows the use of the t-distribution statistics when calculating the UCL, as long as there is no evidence to the contrary. Being able to use the t-distribution makes the mathematics much simpler than might otherwise be. Within this context, "evidence to contrary" is indication that the distribution is not well-behaved, as reflected by a high relative standard deviation (RSD) for three or more ICS replicates. For that reason, using the t-distribution to calculate the 95% UCL is recommended only when the RSD is less than 1.5. For RSDs between 1.5 and 3, the recommendation is to use the nonparametric 95% Chebychev UCL. For RSDs greater than 3, it is recommended to use the nonparametric 99% Chebychev UCL. A spreadsheet programmed to calculate these values from data results is available

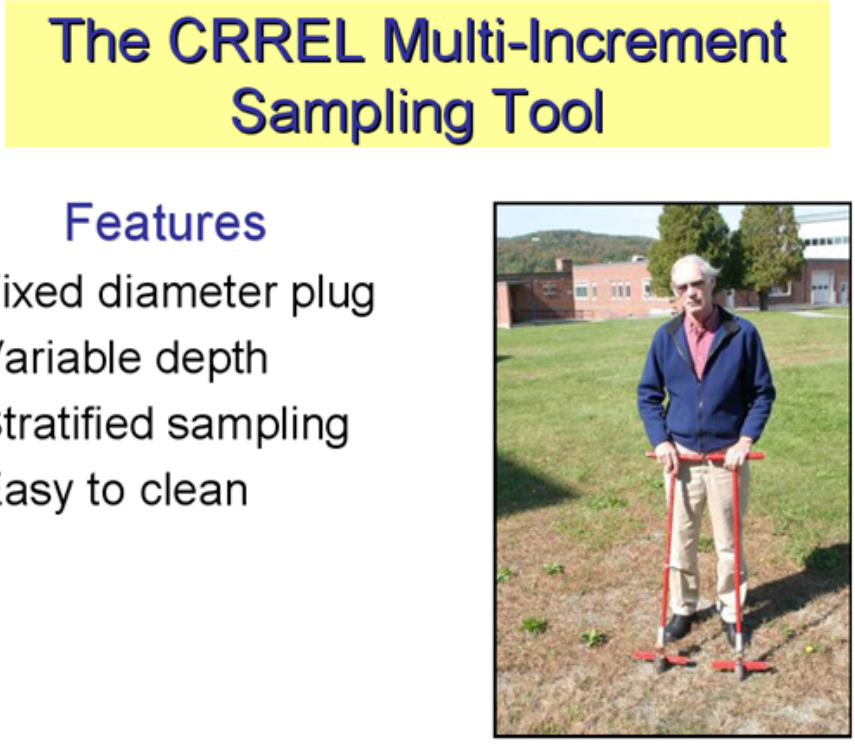




5. Experts feel that 30 increments is sufficient enough to be "safe" for most projects to be used as a default. <u>Be cautious, for there may be sites that do not comply with the assumptions inherent to using statistical techniques. Seek expert advice when in doubt.</u>

Experts also feel the number 30 is low enough to be reasonably practical to implement. Of course, ease of implementation is highly dependent on what sampling and preparation tools are being used. Relying only on conventional tools and procedures can slow the process of collecting increments. A special sampling tool was developed by USACE's Cold Regions Research and Engineering Laboratory (CRREL) to easily collect increment plugs by stepping on it from a standing position. Recall that these increments are immediately pooled together. Therefore, decontaminating the sampling tool (beyond simple wiping) between each increment within the same SU or DU is not necessary because tiny amounts of carry-over are not an issue. This speeds sample collection

considerably. The sampling tool, however, should be properly decontaminated between ICS samples. (Graphic below courtesy of CRREL)



USEPA risk guidance reports that site experience found that 20 to 30 samples did a reasonably good job of estimating the mean. It states:

> "How many samples are necessary to calculate the 95 percent UCL? Sampling data from Superfund sites have shown that data sets with fewer than 10 samples per exposure area provide poor estimates of the mean concentration (i.e., there is a large difference between the sample mean and the 95 percent UCL), while data sets with 10 to 20 samples per EA provide somewhat better estimates of the mean. Data sets with 20 to 30 samples provide a fairly consistent estimate of the mean (i.e., the 95 percent UCL is close to the sample mean). Remember that in general the UCL approaches the true mean as more samples are included in the calculation." (USEPA, 1992a, page 3).

All of the above are contingent upon the size of the area to be represented by the ICS sample. Thirty increments may not be sufficient for a very large area, depending on how the data may be used. The 30 increments should be sampling from the same contaminant population, and as land areas get larger, they inevitably include more than one population. Populations may overlap, or one population may fade out to be replaced by a different population. Either way, including data from two or more different populations in a single data set for these kinds of statistical applications can present difficulties for sampling and data interpretation. On the other hand, the definition of what constitutes a "population" might include areas that have a high degree of variability.

However, when the population is defined as an exposure unit (EU), the EU is typically considered as a single DU. As discussed in USEPA RAGS (1989c):

> "The concentration term in the intake equation is the arithmetic average of the concentration that is contacted over the exposure period. Although this concentration does not reflect the maximum concentration that could contacted at any one time, it is regarded as a reasonable estimate of the concentration likely to be contacted over time. This is because in most situations, assuming long-term contact with the maximum concentration is not reasonable."

RAGS goes on to point out that there are exceptions if the presence of hotspots occurs in locations where higher exposures are possible:

> "If a hotspot is located near an area which, because of site or population characteristics, is visited or used more frequently, exposure to the hotspot should be assessed separately. The area over which the activity is expected to occur should be considered when averaging the monitoring data for a hotspot. For example, averaging soil data over an area the size of a residential backyard (e.g., an eighth of an acre) may be most appropriate for evaluating residential soil pathways."

EPA guidance clearly anticipates that project teams may carefully consider contaminant variability and the exposure scenario as important factors when setting the dimensions of the DU. The sampling unit (SU) concept is also very useful when faced with scenarios where detecting and measuring hotspots are matters of concern.

Scaling up to a larger area might involve increasing the number of increments into the 30 to 60 range. As mentioned before, the "correct" number of increments is determined by the degree of heterogeneity within the DU. Deciding exactly how many increments are needed for larger areas depends on historical information, if available, or on the results of a pilot study that determines the spatial variability for the area and the spatial increment density that sufficiently controls for that variability. If these are not available, choosing the number of increments depends upon the judgment of the technical planning team. If there is considerable uncertainty in selecting the number of increments, ICS replicates should be collected as a quality assurance (QA) measure. If the increment density is not sufficient to support DU decision-making, this may show up mathematically when calculating the RSD for the DU replicates. The greater the variability between the replicates, the higher the UCL will be. If there is high variability (i.e., the RSD is greater than 1.5), the need to use the nonparametric Chebyshev UCLs can greatly increase the UCL. If the UCL and the mean bracket the action level, this is an indication that there is more variability in the data set than can be tolerated in the decision-making process. If there is sufficient control over variability, the mean and the UCL should both fall on the same side of the action level. Note that even high variability in the data (a wide distance between the mean and UCL) might not be important if the field concentrations are far from the action level. If the mean is very low compared to the action level, the UCL should still fall below the action level.

Ideally, the number of increments should be determined during systematic planning, based on by the CSM and the degree of statistical variability present in TEQ concentrations across the SU or

DU to be evaluated. Having relevant historical data is the easiest way to get an estimate of the standard deviation (SD).

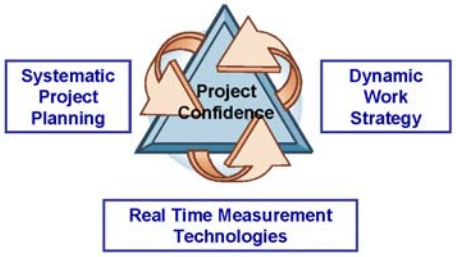
Another option is to obtain an estimate of the field variability from a pilot study either using standard analytical methods or screening analyses. Standard analyses are expensive, but the costs saved because of the information gained from the pilot study may exceed the funding spent on it. There is also the possibility of using screening analyses to measure variability during a pilot study. Techniques considered "screening" for TEQ's include immunoassays and bioassays. Some of these tests can respond to very low (part per trillion) concentrations. Many project teams may not be comfortable using and interpreting the data from screening analyses at this time. Worksheet templates (for UFP-QAPP Worksheets #19, 23-25, and 28) for use with various TEQ screening techniques, and suggestions for their application, may be developed based on DMA and pilot study activities from early applications at assessment sites.

Appendix 3 See also http://www.clu-in.org/download/char/demonstrations_of_methods_applicability.pdf

Demonstrations of Method Applicability under a Triad Approach for Site Assessment and Cleanup — Technology Bulletin
August 2008

Since its inception in 1995, the U.S. Environmental Protection Agency's (EPA) Brownfields Initiative and other revitalization efforts have grown into major national programs that have changed the way contaminated property is perceived, addressed, and managed in the United States. In addition, there has been a shift within EPA and other environmental organizations in the way hazardous waste sites are cleaned up. Increasingly, project managers, regulators, technology providers, and other stakeholders are recognizing the value of implementing a more dynamic and flexible approach to site cleanup that focuses on real-time decision-making in the field to reduce costs, improve decision certainty, and expedite site closeout. The approach, known as Triad, uses (1) systematic project planning, (2) dynamic work strategies, and (3) real-time measurement technologies designed to increase confidence in the project (Figure 1).

Figure 1. The Triad Approach



Triad's best management and technical practices have been successfully implemented in a variety of regulatory frameworks, including Brownfields, Superfund, the Resource Conservation and Recovery Act (RCRA), management of underground storage tanks (UST), and voluntary cleanup programs. As a result, the EPA Brownfields and Land Revitalization Technology Support Center (BTSC) is preparing a series of technical bulletins to provide additional information on implementing specific aspects of the Triad approach. These bulletins are intended for technical project managers and team members. Non-technical managers or stakeholders may also present these bulletins to consultants and service providers to ensure Triad best management and technical practices are implemented appropriately at their sites. These bulletins provide sufficient information for less technical project managers and team members to request that critical Triad project elements be included in scope of work and planning documents.

Demonstrations of Method Applicability (DMA) are a key component of using real-time measurement technologies and are presented in this bulletin through:

1. Answers to frequently asked questions on key aspects of DMAs
2. Examples of DMAs performed at hazardous waste sites:
   - Wenatchee Tree Fruit, Wenatchee, Washington
   - Poudre River, Fort Collins, Colorado
   - Fort Lewis Small Arms Firing Range, Fort Lewis, Washington
3. Sources of additional information for communities and project teams that desire to implement DMAs and the Triad approach.

About the Brownfields and Land Revitalization Technology Support Center (BTSC)

EPA established the BTSC (see *www.brownfieldstsc.org*) to ensure that brownfields and other land revitalization decision-makers are aware of the full range of technologies and technical support services available for site assessments and cleanups and to help them make informed decisions about their sites. The center can help federal, state, local, and tribal officials evaluate strategies to streamline the site assessment and cleanup process at specific sites; identify, review, and communicate information about complex technology options; evaluate contractor capabilities and recommendations; and plan technology demonstrations. BTSC is coordinated through EPA's Office of Superfund Remediation and Technology Innovation (OSRTI) and works through EPA's Office of Research and Development (ORD) laboratories. The center also works closely with EPA's Office of Brownfields Cleanup and Redevelopment and in partnership with the U.S. Army Corps of Engineers (USACE) and Argonne National Laboratory.

Localities can submit requests for assistance through the EPA Regional Brownfields Coordinators, online, or by calling 1-877-838-7220 toll free. For more information about the BTSC, contact Carlos Pachon at (703) 603-9904 or *pachon.carlos@epa.gov*.

What is a DMA?

A DMA is an initial site-specific performance evaluation for a wide range of sampling, testing, and data management tools. It is a concept founded in EPA SW-846 guidance (www.epa.gov/epaoswer/hazwaste/test/sw846.html) and is based on the principles of EPA's performance-based measurement system (PBMS) initiative (www.epa.gov/SW-846/pbms.htm). A DMA usually falls into one of two categories: (1) a comparison of a field-based method with a more established laboratory-based method to demonstrate the usefulness of the field-based method, or (2) a test to evaluate whether a particular tool will work on a specific site. Both types of DMAs may be needed at a single site, and the exact format of the DMA will depend heavily on the site characteristics, the history of investigations at the site, and the intended use of the data.

The DMA serves several different purposes for many applications, including showing whether a technology will be effective at the intended site, but also to optimize how it will be used collaboratively with other information sources at the site. DMA data are of particular importance relative to understanding the potential effects of matrix heterogeneity and sample support on data quality. During the DMA, the types and frequency of quality assurance and quality control (QA/QC) procedures are often tested for adequacy, and preliminary field-based action levels are developed for comparison with site decision criteria. Methods for data sharing and management are also tested to assure a project can proceed in real-time.

In addition, a DMA can be used to evaluate technologies for generating analytical data (or other information) both in the field and in an off-site location that will provide information appropriate for meeting project decision criteria. The ability of technologies such as X-ray fluorescence (XRF), immunoassay (IA), ultraviolet (UV) fluorescence, and direct sensing tools such as the membrane interface probe (MIP) and laser induced fluorescence (LIF) to produce decision quality data has been well documented (www.epa.gov/ORD/SITE and www.epa.gov/etv) and their suitability to a site can be evaluated through a DMA. Extensive literature and performance data are available for some technologies, and these data should be reviewed by project teams before a DMA is designed.

A DMA can also provide information on cost and performance that can be used to optimize collaborative data collection using technologies for generating analytical data (or other information) both in the field and in an off-site location. Additionally, a DMA can offer stakeholders an understanding of the site-specific performance of a technology while at the same time it provides the basis to optimize standard operating procedures (SOPs) for deployment.

DMAs are performed easily and affordably before mobilization, or as an early component of a field program. Advice on the specific technology and assistance to set up a DMA sometimes are available from equipment vendors or service providers.

The question often arises: "How do DMAs fit into the data quality objective (DQO) process?" Various aspects of DMA planning and implementation fall under a number of DQO steps, but the most important one is Step 7: Select the most resource-effective sampling and analysis strategy that meets the performance criteria.

When is it necessary to perform a DMA?

DMAs may be used when the project team works with a technique that previously has not been used at the site. Site-specific factors often may render an otherwise useful technique unsuitable and can result in high and unnecessary project costs if they are not discovered early in the project. A DMA can quickly ascertain whether the new technique is suitable for use at the site, allowing the project team to identify an alternative if the proposed technique is not suitable. Conversely, the project team can proceed with confidence, realizing the benefits offered by the technique, if the DMA suggests that it will be effective.

A DMA may be necessary when:

- A project will depend heavily on field-based results to make real-time decisions.
- Experience indicates a technology's performance is variable from site to site.
- Heterogeneity and the cost of cleanup are high.
- The chemistry of contaminants is complex.
- A specific relationship is needed between collaborative forms of data sets to support decision-making.
- Stakeholder acceptance requires that the utility of a technology or approach be evaluated.

What are the benefits of performing a DMA?

A well-planned DMA can simultaneously test, refine, and coordinate many project design parameters before full-

scale project activities are under way. Project design parameters often evaluated during a DMA include sampling and analytical methods, QA/QC procedures, data management, communication and data sharing strategies, collaborative or comparative data needs (for example, technologies matched with other field tools or standard fixed-laboratory analytical methods), project staffing, and the overall flow or sequencing of field activities. A carefully considered DMA can help an entire project run faster and more smoothly, resulting in lower costs, and assuring that the data collected will be adequate for the intended end use. Both field-based and fixed laboratory methods have limitations, and the project team should verify their performance during project startup to avoid generating data or using equipment that does not meet project requirements for precision, accuracy, representativeness, completeness, comparability, specificity, sensitivity, ruggedness, and reliability.

Project-specific DMAs guide the project team in selecting and optimizing collaborative methods and assuring adequate method performance for site conditions and decision criteria. DMA results can be used to develop project-specific action levels; analytical, sampling, and data analysis procedures; QA/QC requirements; and additional data requirements to assure the quality of the decision. Furthermore, A DMA can help set acceptable levels of uncertainty relative to decision thresholds used in the field as part of a dynamic field decision strategy. The fast pace of real-time cleanup projects makes DMAs essential to avoid down time related to problems that arise from inadequate planning for sampling and analysis.

Important functions of a DMA include:

- Providing assurance that the proposed site characterization methods are suitable for the specific project.
- Generating data of known quality.
- Developing initial relationships between field methods or tools and other collaborative methods such as fixed laboratory. These relationships are used to design and focus the QC program.
- Testing a preliminary CSM to refine sampling protocols if assumptions are found to be incorrect.
- Setting preliminary field-based action levels to be used for real-time decision-making.
- Establishing the readiness of field personnel, equipment, and procedures before full-scale work begins at a site.
- Assessing alternative strategies as contingencies should the performance of the intended methods be compromised by unanticipated problems.

As noted earlier, DMAs can be used to test the suitability of technologies for generating any data, whether the analysis occurs in the field or in a traditional laboratory. Chemical data-oriented DMA tasks involve collecting, preparing, and analyzing samples from a site-specific matrix (soil, water, air, and tissue, for example).

DMAs also serve important non-analytical functions and can help evaluate whether a project is ready to proceed. DMAs can be used to test the preparedness of field personnel and service providers, as well as to evaluate the adequacy of logistical and data management plans. For example, sample throughput and analysis times can be more accurately estimated. Likewise, materials and personnel needs can be balanced and documentation procedures clarified. In addition, instrument compatibility and data exchange or upload protocols can be verified and debugged as necessary. Assessing logistical feasibility in this manner is especially important when the project team uses dynamic work strategies. This aspect of a DMA can help evaluate practical constraints for work at the site in relation to the timing of sample collection and analytical throughput, including field analytical equipment, labor, sample storage, and the cost and supply of consumables.

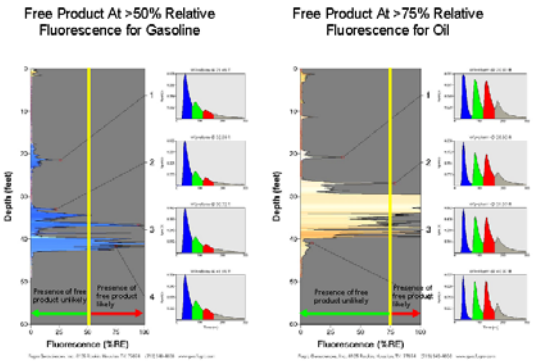
The DMA can also be used to "test-drive" real-time decision support tools (DSTs). These tools include electronic data management procedures, global positioning system (GPS) and surveying equipment, and modeling, mapping, and data display software. This aspect of a DMA will improve the ease of use during full-scale field mobilization by ensuring operators can do the job and identifying those aspects of a DST that can be improved. In some cases, the project team may decide a different DST is more appropriate for one particular portion of a project versus another.

Finally, the DMA can be used to assess the appropriateness and performance of proposed generation techniques for data other than chemical, such as geophysical, geotechnical, or direct sensing or probing methods. The presence of site-specific interferences that could compromise the performance of these tools can be tested as a result. Interferences for geophysical techniques could include tree leaf cover, seasonal wetland

areas, low power lines, fences, shallow bedrock, salts, and interlocking sands. Mineralogical interferences or other geological conditions can also affect the performance of direct-sensing equipment such as LIF probes. The presence of some forms of calcite or specific clay materials also can impair the utility of LIF. If interferences are identified during the DMA, alternative strategies for dealing with them can be developed before full-scale work is undertaken at the site.

DMAs for direct-sensing tools include many of the same techniques used for evaluating other field analytical tools. They can involve building relationships between sensor response and analytical data or other forms of comparable information such as visible staining, free product, soil saturation, or physical characteristics of the matrix. Figure 2 provides an example of developing relationships between LIF response and visual core observations. At this site, relative fluorescence for various product types such as gasoline, diesel, and oil were used to estimate the presence or absence of free product. When used with LIF logs to estimate product thickness, these values allowed members of the technical project team to estimate contaminant mass and optimize locations and depths for a product removal system.

Figure 2: Correlations of LIF response and presence of free product.



What are key concerns to address in designing a DMA?

At least four aspects of a data-focused DMA should be considered during its design, including the following:

- What are critical aspects of the preliminary conceptual site model (CSM) that should be tested to assure project success? These aspects may include assumptions about the locations and nature of suspected releases, the degree of matrix heterogeneity, and the impact of sample processing bias. Understanding matrix heterogeneity is critical in evaluating the number of samples required for statistically based sampling designs. Without some information about the specific site and performance of the analytical method, the appropriate number of samples needed to achieve a desired level of decision confidence cannot be correctly identified.
- Is it important to evaluate the cost-effectiveness and bias of multi-increment sampling? It may be important to compare the cost-effectiveness and bias of traditional grab sample collection and mathematical averaging procedures against a multi-increment (physical averaging) sampling strategy.
- Do changes in sample support (the size, shape, and orientation of a sample) dramatically affect analytical data results? Is the level of effort associated with various sample preparation and analysis techniques worth the benefit of higher precision, accuracy, or control of bias?
- Are project decisions of a qualitative or quantitative nature? In some cases, a "yes or no" answer is all that is needed, while for others a more quantitative result is needed. For example, in some cases the decision is only whether free product is present or absent in the subsurface. In contrast, risk estimation often requires data in the form of quantitative concentration results.

DMAs should be designed to address those issues that most often provide the greatest source of uncertainty: sample heterogeneity and short-scale spatial variability. The resulting mismatch between the volume of the sample analyzed and where it is collected versus the data result that will be extrapolated to a significantly larger volume of material can be significant. Therefore, the DMA should appraise sampling designs (such as multi-increment designs), sample collection techniques (such as low-flow purging of ground water wells versus passive diffusion samplers), and sample preparation procedures (such as in situ versus ex situ readings and options for sample homogenization or fractionization based on soil properties such as particle size).

A DMA often involves "split samples" that are carefully prepared to minimize matrix heterogeneity and analyzed by two or more different techniques to establish relationships. Parametric and non-parametric techniques are commonly used to evaluate these relationships and establish decision quality for collaborative data sets.

Parametric statistical methods use assumptions about the data's underlying statistical distribution shape (normal, lognormal, or other). If those assumptions are invalid, the statistical conclusions may not be reliable. Non-parametric techniques do not require that as many assumptions be true, so they are more broadly applicable to the properties of environmental data.

Comparability is quantified by establishing the frequency that results from different techniques agree with each other with respect to a declared reference point. Different points of reference can be used, but the most common strategy in Triad projects is to establish comparability in terms of the decision being made on the data. These data may require quantitative comparability (such as if or when two data sets are combined to calculate risk assessment parameters) or qualitative comparability for agreement at the compliant or non-compliant decision threshold.

A comparability DMA can be used to demonstrate the suitability of field-based technologies or project-specific modifications to improve the performance of an established fixed-laboratory method. The techniques to be compared include sampling related methods as well as analytical methods. Understanding the effects of sample heterogeneity is extremely important when samples will be split for the different analytical methods to be evaluated. A valid comparison of data sets requires thorough homogenization of samples before they are split to ensure both methods see the same sample characteristics that will be used to make a decision. To understand method differences, known or blind QA samples (spikes, replicates, reference materials, or blanks) are also often subjected to comparative analysis to assure technical team members that both methods are providing representative results.

What data deficiencies can be addressed using a DMA?

Site-specific method reporting limits (MRLs), precision, bias, false positive rates, and false negative rates can be assessed through the DMA process. For example, MRLs and sample reporting limits can be tested by analyzing samples spiked with known amounts of target contaminants and comparing site-specific matrices to find the lowest concentration that can be reliably detected and quantified.

Data from laboratory reference methods and from a field-based method should be compared to see whether they produce data that lead to the same project decision based on established field-based action levels or decision rules.

The ability of two methods to agree for decision-making is an important parameter to examine when comparing analytical methods, especially when methods with lower analytical performance such as immunoassay methods (which measure several closely related analytes and report a single result for the group) are being compared with methods with higher performance, such as gas chromatography (GC) and mass spectrometry (MS) (which are usually able to distinguish between closely related analytes and measure each). In this way, the DMA is a critical component in Triad's efforts to manage the analytical contribution to decision uncertainty.

How can DMA use weight of evidence and collaborative data sets?

Terms such as "weight of evidence" or "multiple lines of evidence" and "collaborative data sets" have been developed to describe these layered data sets. From a Triad perspective, there is a distinction between the two. "Weight (or lines) of evidence" refers to combining information from various different sources into a holistic picture (that is, a CSM). For example, historical information may be used in conjunction with geological, hydrogeological, chemical, and geophysical data to predict contaminant fate and transport.

On the other hand, "collaborative data sets" or "collaborative methods" refer specifically to the strategy of using two (or more) analytical methods to measure the "same" analyte or a surrogate of an analyte. For example, total uranium can be measured by XRF, gamma spectroscopy, and alpha spectroscopy. Collaborative methods are paired so that the strengths of one method can compensate for the limitations of the other. Frequently, a field method is selected for its ability to provide a much higher density of data points than an expensive laboratory method. However, the laboratory method will generally achieve better detection limits and accuracy than the field method. A DMA should be designed to guide the "marriage" of the techniques to produce reliable information that is not biased by the effects of heterogeneity or analytical inaccuracy.

Additionally, alternative analytical methods, particularly any that provide results in "real time," can be used to optimize the decision making process. For example, the real-time decisions and high data density possible with field methods can reduce the volume of material removed during cleanup by more precisely defining and confirming the actual contamination footprint. Real-time data can in this way improve confidence in the decision and limit "surprises" after a project is complete.

What is involved in performing a DMA?

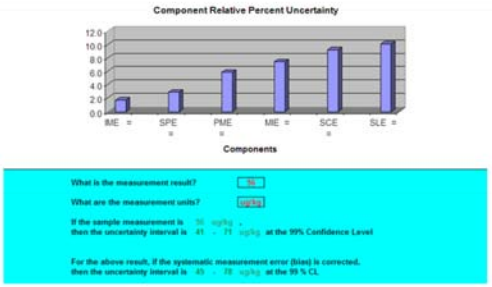
Designing an appropriate DMA is specific to the technology or technique being employed and to the project, the site matrix or matrices, the effects of heterogeneity, sample support (the size, shape, and orientation of a sample), the expected use of the data, and a myriad of other factors. During systematic planning, the project team may evaluate potential candidate technologies or strategies for use at the site. Technologies and strategies that can improve project efficiency and the CSM, increase data density, and reduce uncertainties associated with decision making are most often targeted. Project teams are encouraged to employ the services of an experienced Triad practitioner when technologies or strategies are short listed. Although there is no generic format for designing a DMA, a number of activities are often involved.

- Evaluating the strengths and limitations of technologies or techniques to be used on site samples.
- Evaluating sample support, throughput, ease of use, manipulation and storage of data, and other logistics so that the process is optimized.
- Collecting and analyzing QC samples to evaluate the uncertainties that are the largest contributors to total measurement error. Project resources can then be allocated to control for those activities with the greatest effect (Figures 3 and 4).

Figure 3. Uncertainty Sources and Associated QC Samples

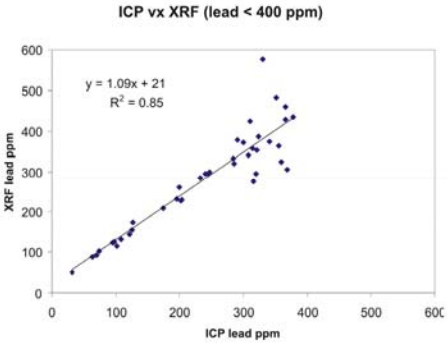
Uncertainty Sources	Source Symbol	Analytical Sample	Analytical Sample Symbol
Intrinsic (Instrumental) Measurement Effects	IME	Instrument Calibration Standard	ICS
Spike Preparation Effects	SPE	Initial Calibration Verification Standard	ICV
Preparation Method Effects	PME	Laboratory Control Sample	LCS
Matrix Interference Effects	MIE	Matrix Interference Sample Matrix Spike/ Duplicate Sample	MIS MS/MSD
Sample Collection Effects	SCE	Field Replicate (Duplicate) Sample (Collected from same location and during same sampling event time)	FSR
Sample Location Effects	SLE	Co-Located (Same Location) Sample (Collected 0.5 – 3 feet away from field sample)	CLR
Sampling Site Population Effects	SSE	Site field sample collected from the environmental site for the study	SFS

Figure 4. Output from an Uncertainty Evaluation



- Collecting information to establish initial relationships with data from the fixed laboratory or other collaborative information. The collaborative relationship (data comparability) can be evaluated using a variety of options.
- Completing parametric statistical techniques, such as linear regression (Figure 5). Although commonly applied, caution should be used with linear regression. The correlation coefficient (R^2) is universally used as a measure of a good relationship between two methods, but can be misleading. Examining the slope and y-intercept can be far more informative and less distorted by isolated high values.

Figure 5. Sample Linear Regression



- Using non-parametric techniques (often more useful for establishing comparability). These techniques are "common sense," but still powerful aids to decision making. Examples include scatter plots, calculating decision error rates and establishing investigation levels to use with the alternative technique (Figure 6).

Figure 6. Sample Investigation Levels

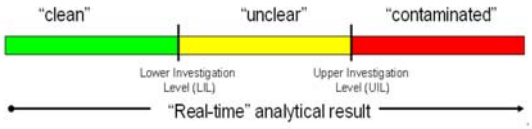
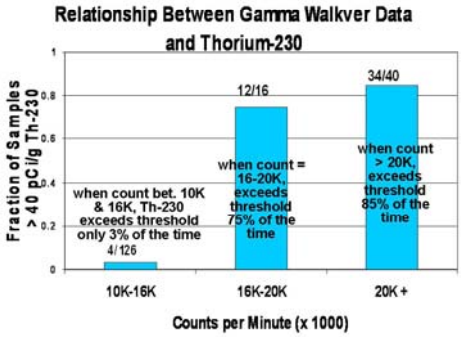


Figure 7. Typical Outputs of Investigation Levels and Decisions Bins Using Non-Parametric Methods



- Using "decision bins" to establish investigation levels can guide confident decisions made on field data. (Figure 7).

- Identifying potential interferences, bias, false positive and false negative rates, and other issues. Depending on project timelines, the Quality Assurance Project Plan (QAPP) can be formally or informally updated and optimized as a result of the DMA to manage QA issues and produce data of known quality. The plan should also identify steps to address violations of QC criteria should they occur during the full-scale field effort.

- Using data collected during the DMA as the input values to construct a statistical sampling design. One of the acknowledged pitfalls associated with using classical statistical tools in sampling design is that project teams seldom have a sound estimate of total measurement error to use in establishing sample quantities, grid sizes, and other factors. With results from a DMA, project teams can use classical statistical tools (such as the Visual Sample Plan software, http://vsp.pnl.gov/) more effectively in sampling design because they have generated site-specific information on method error.

- Evaluating site-specific method error helps establish initial collaborative relationships that can be refined as the program progresses. These relationships provide a framework for indicating problematic samples or "out of control" QC issues.

- Providing insight into how the full set of data may be statistically evaluated. Statistical methods such as those described in the guidance on Data Quality Assessment (EPA 2006b) may be examined for effectiveness and used to test basic project data assumptions, contaminant distributions, and sampling designs planned for use at a site.

- Testing the suitability of data visualization and management strategies.

How are results of an analytical DMA applied?

If the DMA is properly designed, the data can be used for the following:

- Support development or refinement of the CSM
- Estimate matrix and contaminant concentration variability at different spatial scales
- Identify potential interferences
- Ascertain whether particle size is correlated with contaminant concentrations
- Evaluate the value and effectiveness of different sample collection and processing techniques to optimize SOPs
- Establish a comparability relationship between two measurement systems
- Establish proper decision logic and sequencing of data collection

In this section, readers will be introduced to some of the basics of environmental decision making. Understanding the context in which decisions are made is essential for discerning how managing associated uncertainties affects the decisions. Managing decision uncertainties is important to developing realistic and protective field-based action levels for a site. Overly protective standards can significantly increase project costs, while less stringent standards may lead to controversy or surprises later in the project.

Under Triad, use of collaborative data becomes essential to provide sufficient density, and these data sets are then compared to improve decision certainty. A variety of methods can be used to assess the comparability between measurement systems. "Measurement system"

refers to the combination of techniques used to collect and process a sample in addition to the actual analysis. For example, analysis of split samples that have been well-homogenized (which, depending on the matrix and analyte, may require grinding to a uniform particle size), will be only part of the measurement system compared. The overall measurement system for both XRF and the compared laboratory technique must be considered for technologies such as XRF spectrometers that can be used for in situ sampling and where the effect of sample properties cannot be removed from the analytical process.

Although there are a variety of mechanisms to assess the comparability of data sets from two different measurement methods, traditional statistical techniques are widely used to begin the process of exploring the data. Traditional statistical evaluations of data sets can include summary statistics and statistical plots (box and whisker plots, histograms, and probability plots) to evaluate distributional characteristics of the sample population (normal, lognormal, or other) that decide what types of statistical manipulations are warranted.

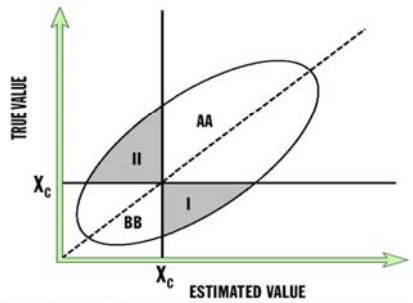
Field-based and fixed-laboratory results can be compared by developing correlation scatter plots, or by calculating best-fit lines and correlation coefficients to describe the mathematical relationship between the data sets. If a field method is shown to be biased high uniformly across a site, the bias might be used to provide a natural safety factor when compared with regulatory limits. Alternatively, if the bias is highly consistent and predictable across samples and concentrations, adjustment for the bias is a possibility.

One common use of a DMA is to demonstrate that a less-rigorous analytical method correlates well with an established method. Best-fit lines and correlation coefficients may be used for this purpose. If correlation is lower than expected, a DMA may show that decision error is low enough that the less-rigorous method is still acceptable for the purpose of increasing sampling density to delineate contaminant footprints and control data variability on large and small spatial scales. This ability is valuable since matrix heterogeneity and small-scale variability are often the largest contributor to total measurement error in environmental data.

In real life, the "true value" is unknown. Any data result, no matter how good the analytical method, is an approximation of the true value. *If sample heterogeneity and interferences are controlled*, the more sophisticated analytical method will be closer to the "truth." However, the expense of these methods can limit the number of samples that can be analyzed. Non-representative data that are biased by matrix problems can lead an unwary decision maker into costly decision errors. Using two analytical methods helps ensure that matrix heterogeneity does not mislead the decision maker. One analytical method (usually the fixed-laboratory method) will produce more accurate measures of concentration, and that method is used as the surrogate for the "true value." The other method (usually a field method) produces a more accurate representation of sample representativeness, but at the expense of data accuracy. Therefore, in a DMA, traditional laboratory results are assigned as the "true value" (y axis) in Figures 8 through 10. Field results are assigned as the "estimated value" (x axis). "Investigation levels" are selected through the process of comparing the two sets of data and minimizing the likelihood of decision errors. Field results that fall above and below the upper and lower investigation levels can support confident decisions. Field results that fall within the concentration range between the investigation levels require analysis by the more accurate laboratory method.

Figure 8 illustrates two types of decision errors possible when two sample analysis methods are compared. Assume all data points fall within the ellipse. "False positive" decision errors (also called "false dirty" decision errors) occur when a data result falls above an action level when the true result is below the action level, and the decision maker undertakes unnecessary remediation. "False negative" decision errors (also called "false clean" decision errors), occur when a data result is below the action level when the true result is actually above the action level. If the decision maker accepts the data at face value, erroneous decisions are possible that potentially increase risk to human or ecological receptors.

Figure 8. Misclassification Ellipse

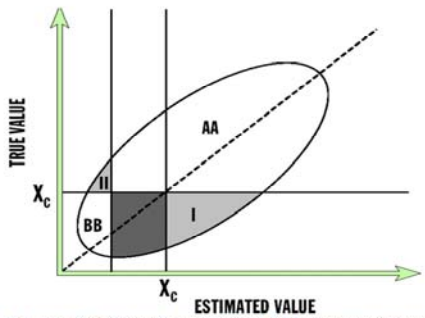


Xc denotes the action level, areas I and II indicate the false positive and the false negative decision error zone, respectively. Areas AA and BB indicate zones of consistent decisions between the data results and the true values.

Figure 9 shows how using a safety factor below the action level can reduce the false clean (false negative) decision error rate.

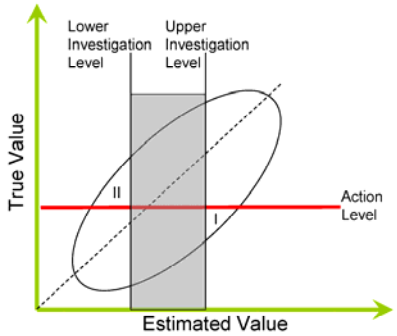
Figure 10 shows how safety factors both above and below the action level reduces both the false clean and false dirty decision errors.

Figure 9. Misclassification Ellipse with Safety Factor.



Xc denotes the action level, area I indicates the false positive error zone, and area II indicates the false negative error zone that has been reduced by the safety factor. Areas AA and BB indicate zones of consistent decisions between the data results and the true values.

Figure 10. Multiple Investigation Levels.



Both types of errors are reduced. Shaded area represents data results falling between the investigation levels that require more accurate analysis.

What else should project managers and technical staff consider when planning a DMA?

Refining decision criteria and decision logic based on the results of a DMA can significantly improve project performance. Results from the DMA should be integrated as quickly as possible into work strategies to assure project efficiency (See EPA 542-F-05-008, "Use of Dynamic Work Strategies Under a Triad Approach for Site Assessment and Cleanup — Technology Bulletin, www.brownfieldstsc.org/pdfs/DWSBulletin.pdf).

Project managers must identify resource needs to support real-time decision-making during the DMA. These resources include DSTs and associated expertise. For example, a DST may be required to assist in developing and verifying field-based action levels.

Project managers and technical staff should refine the type and level of field documentation required based on the DMA. Site-specific work plans and SOPs for field methods with sufficient flexibility can be easily revised as more is learned about a site, even after the DMA is complete. Team member responsibilities should be consistent and modified as needed based on the DMA.

Creativity and flexibility in procurement and contracting is often needed for a DMA, or in response to a DMA. Review EPA's procurement guide at www.brownfieldstsc.org/pdfs/procurement.pdf as a starting point for procurement and contracting strategies for Triad investigations. Unitizing or classifying costs per analytical sample or borehole, for example, is an illustration of a financial strategy that allows project planners to accurately track costs in real time as a dynamic investigation progresses. Potential vendors may provide free resources for the DMA to market and demonstrate the applicability of the technology, reducing the cost to the project — most commonly with newer and relatively unproven technologies.

Workload balancing and task sequencing are examples of strategies used to ensure that project team members are aware of the project's time-critical tasks. Team members should work together to prioritize each task so that no task slows the entire effort while others (drillers, samplers, or analytical chemists) are idled, but still billing time. A DMA can provide important information on potential bottlenecks in data generation and flow, so that effective field coordination and sequencing strategies can be developed for the main field investigation.

How is a DMA documented?

A DMA is documented through a variety of formal and informal means. Project plans such as Sampling and Analysis Plans (SAPs) and QAPPs are formal documents that undergo mandatory review. When they are written for a Triad investigation, these documents outline the DMA to refine the data collection schemes and strategies to manage uncertainty. Site-specific SOPs for field methods are useful for documenting the outcome of a DMA in the

form of exact procedures to be used at a given site, especially when there are deviations from vendor-recommended procedures or published methods (such as SW-846 methods).

More informal methods, such as memoranda of understanding, meeting notes, project Web sites, E-rooms, and electronic bulletin boards, also document the DMA process. These informal methods are particularly useful to document stakeholder participation and buy-in for Triad investigations. Informal discussions with stakeholders after the DMA can also be useful to accelerate document comment, revision, and submittal.

Regardless of the method used to document a DMA, good records are essential to scientifically validating and legally defending selection and use of analytical methods and, ultimately, the conclusions based on data generated in the field. Project teams are strongly encouraged to plan for and complete DMA documentation.

Are DMAs difficult, time-consuming, or costly?

Generally, DMAs can be completed with 20 or fewer samples, but the level of effort can be scaled to the magnitude of expected site work. However, more data can be extremely valuable in the case of many real-time technologies where costs per sample are relatively low. If linear regressions will be generated, a good rule of thumb is at least 10 paired samples; however, 20 or more can provide exceedingly robust statistical evaluations. A key concept for analytical DMAs is to focus sample pairs around action or decision points (for example, 5 low values, 5 higher values, and 10 in areas around action levels). Using real-time measurements provides a level of assurance that samples submitted for fixed-laboratory comparative analysis are in the range of interest. Data sets with high percentages of non-detected pairs are not beneficial for statistical evaluations.

Information collected during the DMA will provide a basis for establishing QA/QC protocol, sample support, preliminary relationships for collaborative data sets, load balancing, and sequencing field activities. The DMA results are a means for optimizing use of resources and become part of the final data set that will be compiled to reach project decision points.

The cost and time required for a well-designed DMA are usually a small fraction of the cost of a full-scale field program. When the project team designs a DMA, the cost and time allotted should be proportionate to the impact of the DMA on reduction in uncertainty about a site condition. It is expected that relationships evaluated under the DMA will continue to be refined as more data become available. Another cost savings consideration is the use of archived material (where appropriate) for comparative analyses that may have already been completed. These samples provide the advantage that concentration ranges of contaminants of concern (COCs) to target samples in the primary areas of interest will be known.

In any project, the methods being used will be under scrutiny. When a DMA is not conducted on a limited number of samples, the data collected during the full site investigation must be used to demonstrate method performance. The DMA therefore provides an opportunity to change tactics affordably if a method does not perform as expected, compared with the alternative of having to change tactics after the full site investigation.

Appropriate professional expertise and good communication among team members about their data needs are critical to planning a successful DMA. The cost of a DMA is recouped many times over though cost avoidance of unusable data and recharacterization efforts.

Are DMAs appropriate in all cases?

Some investigation and cleanups may prescribe set sample numbers or recommend limited sampling through guidance. In these cases, DMAs may not be appropriate. Similarly, resources may not be adequate at some sites with limited grant funding to accommodate a DMA. Projects with adequate resources to employ established mobile or fixed-laboratory methods at sufficient density may be inappropriate, while those that require method modifications or careful examination of sampling and spatial uncertainties may benefit significantly from DMAs. Even if only fixed-laboratory methods are used, a DMA may be considered if there is any question about matrix interference effects. A relatively limited number of pilot samples can save large sums of money by detecting extraction issues and interference problems at the start of a program.

It should be noted, however, that a DMA is beneficial for most applications precisely because a particular field analytical technique, direct-sensing tool, or innovative strategy is identified as potentially applicable to cost-effectively increase data density, refine the CSM, or address small-scale variability and matrix heterogeneity. In some cases, sampling locations for an early assessment are obvious (for example, areas of visible

staining, product, lagoons, or discharge points), while other cases are more complicated, and identifying appropriate sampling locations is a problem. Depending on the nature of the suspected contamination, some sample material can be archived and potentially used later as part of a DMA for an expanded assessment or additional investigation.

Adding limited additional cost for analyzing field analytical methods, direct-sensing tools, or other innovative technologies does not significantly raise project expenses at sites with elevated expenditures associated with collecting deeper subsurface samples. In these cases, inexpensive analytics and direct sensing tools can provide greater vertical density to help target locations for more expensive traditional laboratory samples. Furthermore, the increased density can support a more efficient design of cleanup strategies when required, leading to lower project lifecycle costs.

Finally, the definition of a Brownfield — "a property, redevelopment, or reuse which may be complicated by the presence or potential presence, of a hazardous substance, pollutant, or contaminant" — underscores the need for higher data density and collaborative data sets that accompany DMAs. Regardless of whether significant contamination or just the perception of contamination is present at a property, DMAs used to optimize sampling schemes with innovative tools provide a higher data density that facilitates timely revitalization. These well-designed data sets are particularly helpful to address stakeholder concerns and provide a level of comfort that allows developers, insurance partners, risk partners, public stakeholders, state agencies, local agencies, and others to be involved, invested, and reassured with a project outcome.

EXAMPLES OF DMA IMPLEMENTATION

Example #1: Immunoassay: Wenatchee Tree Fruit Test Plot Site, Wenatchee, Washington

The Wenatchee Tree Fruit Test Plot area contained soil contaminated with pesticide compounds from agricultural research conducted from 1966 through the mid-1980s. The U.S. Public Heath Service (PHS) and EPA used the test plot area to evaluate the effectiveness of various land disposal methods for pesticides. In 1997, USACE conducted an integrated site characterization and remediation program that allowed characterization, excavation, and segregation of contaminated material in real time. Work was completed under a voluntary cleanup program with regulatory oversight of the Washington Department of Ecology.

A DMA was conducted to provide critical input to the project design because the project would use IA methods to drive the dynamic work strategy. The DMA was structured to evaluate both the utility of the IA kits and to develop field-based action levels.

Site Facts

- ✓ Disposal area of an agricultural research facility.
- ✓ Reuse scenarios not identified. Changing land use nearby increased concern that the area should undergo investigation and remediation.
- ✓ Principal threats included off-site migration, contamination of other media, and direct contact.
- ✓ COCs included organochlorine, organophosphorous, and other pesticide compounds.

Highlights of the DMA

The DMA confirmed that the IA test kits were intentionally biased 100 percent high by the manufacturer to reduce the chance of false negative results. The DMA accommodated the response of the kits to structurally similar compounds beyond the target compounds. Taking into consideration the high bias and correlations with fixed-laboratory results, the DMA showed that the DDT test kit result exceeding 5 parts per million (ppm) could indicate elevated levels of DDT, DDE, or DDD. Likewise a cyclodiene kit response of 0.1 ppm indicated the possibility that regulatory action levels for endrin or dieldrin were exceeded. Therefore, these values were selected as the investigation levels to make decisions based on the kit results.

Several modifications to the IA kit procedures were made based on DMA results. For one, pure methanol was used instead of a water-methanol mix, and extraction volumes were doubled to 20 milliliters (mL) to bracket action levels based on cross reactivity and sensitivity results. The

resulting 20-milliliter (mL) extracts were sufficient to run both the DDT and cyclodienes IA analyses.

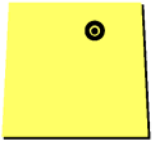
Some fixed-laboratory detection methods for collaborative data were also modified. The project team used MS detection instead of the method specified nitrogen and phosphorus detector (NPD) to improve selectivity and meet project required quantitation limits for the organophosphorus pesticides. Similarly, the team used a GC NPD method instead of high-pressure liquid chromatography (HPLC) for the carbamate pesticides to reduce interference. In addition, the surrogate compound was changed to a compound rarely used in agricultural applications. Non-target compounds and tentatively identified compounds (TICs) from fixed-laboratory methods also became crucial to understanding IA kit response from a broad range of contaminants.

During the characterization phase, the project team continued to collect a subset of samples for fixed-laboratory analysis. These results, used in conjunction with DMA data, indicated that the 5 ppm investigation level for the DDT IA kit was overly conservative. With regulator approval, the DDT IA investigation level was raised to 10 ppm to complete excavation and waste segregation.

No false negative decision errors for the action levels for individual pesticides were encountered. A low percentage of false positive errors (usually associated with of the presence of endosulfans in the samples) was encountered for the cyclodiene kit. Use of the DMA and Triad principles resulted in an estimated savings of 50 percent for total project costs.

More information on the DMA and Triad work conducted for this site is available at the Triad Resource Center Web page: *www.triadcentral.org*.

Example #2: Poudre River Site, Fort Collins, Colorado



The Poudre River Site is located in Fort Collins, Colorado, along the Cache La Poudre River. The presence of coal tar in the river and fuel-related ground water contamination prompted EPA to initiate a Targeted Brownfields Assessment in May 2003. Two DMAs were conducted as part of the Targeted Brownfields Assessment.

Site Facts

- ✓ Site includes a former manufactured gas plant (MGP) that operated from approximately 1900 to 1930.
- ✓ Site includes a former municipal burn landfill that operated from the late 1930s to the early 1960s.
- ✓ Proposed reuse was recreational, commercial, and industrial.
- ✓ COCs included chlorinated solvents and petroleum-related substances.

Highlights of the DMAs

One DMA focused on demonstrating the capability of passive soil gas samplers from EMFLUX (now known as Beacon Environmental) to detect volatile organic compounds in the subsurface.

A full-scale soil gas survey was implemented using 333 devices after the passive soil gas DMA successfully demonstrated the use of the EMFLUX passive soil gas samplers. The data from the study were used to create isoconcentration maps for target analytes, helping to refine the CSM and optimize the field investigation drilling program.

A DMA was also performed for a modified EPA SW-846 Method 8260 used for the analysis of ground water samples on site via a mobile laboratory. The project team used the results from the ground water DMA to set the applicable detection and reporting limits for GC/MS results, design appropriate initial calibration and QC protocols, and evaluate the types and concentrations of contaminants expected in ground water at the site. The DMA also provided site-specific information about the accuracy and precision of the method.

Another aspect of the Poudre River Site field program that showed the usefulness of a DMA study was geophysical survey work. A ground penetrating radar (GPR) survey conducted at the site suffered from poor signal penetration because of soil conditions. However, the performance-based contract used for the work did not allow the cost of this GPR work to be billed against the program. Had the GPR vendor conducted a DMA, this problem would likely

have been discovered earlier and at significantly lower cost to the vendor.

More information on the DMA and Triad work conducted for this site is available at the Triad Resource Center Web page: www.triadcentral.org.

Example #3: X-ray Fluorescence: Fort Lewis Small Arms Firing Range, Fort Lewis, Washington

In 2003, USACE used the Triad approach to expedite site characterization and remediation of contaminated soil at the former Evergreen Infiltration Training Range in Fort Lewis Washington. A dynamic sampling and analytical strategy based on rapid field-based analytical methods was used to streamline site activities and save resources while increasing confidence in remediation decisions.

Initial evaluations included a suite of metals associated with small arms firing ranges (antimony, arsenic, copper, iron, lead, tin, and zinc). The DMA indicated that lead was the primary risk driver given regulatory thresholds and site action levels. After the DMA, the remaining characterization and remediation work at the Fort Lewis site focused on lead.

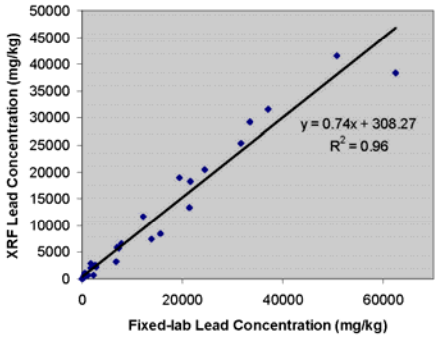
Site Facts

- ✓ Site is a former small arms firing range that operated over an unknown time period between 1917 and 1965.
- ✓ Proposed reuse was for military barracks.
- ✓ Principal threat is direct contact with contaminated soil.
- ✓ COCs were metals including lead and arsenic.

Highlights of the DMA

At the beginning of the field investigation, a DMA was conducted using field-portable XRF and fixed-laboratory methodologies (EPA SW-846; sections 6010 and 6020). Forty samples were collected and analyzed by both methods. The DMA established a strong correlation between XRF and laboratory data for lead (see Figure 11), even with minimal soil sample homogenization. The measured correlation coefficient (R^2) between the methods was 0.96; however, inspection of the slope and y-intercept indicate some loss of linearity. Examining concentrations for individual sample pairs indicates that XRF results tend to under-report concentrations as concentrations increase above percent levels (10,000 ppm). Under-reporting was not a concern for the project team, however, since various action levels for lead were all less than 1,000 ppm.

Figure 11. Correlation Curve.



XRF data are plotted against data from EPA methods 6010/6020 for Fort Lewis Small Arms Firing Range DMA.

A regression was also generated using results for a subset of split samples in the 0 to 1,000 ppm range for lead to evaluate XRF performance in this critical area. The DMA confirmed that the XRF reliably quantified lead concentrations down to 45 milligrams per kilogram (mg/kg), and so was accurate in locating both "clean" and "dirty" areas. Through the DMA, it was assured that data of "known and documented" quality could be produced. The level of data quality was shown to be sufficient for the project's decisions. Although more intensive sample preparation or use of substitution methods for non-detected XRF values may have produced a better regression, the project's data needs did not require the additional precision.

More information on the DMA and Triad work conducted for this site is available at the Triad Resource Center Web page: www.triadcentral.org.

SOURCES OF ADDITIONAL INFORMATION

The Triad approach is encountering ever greater acceptance by regulatory agencies, as well as by professional and industry organizations. Communities and project teams interested in implementing the Triad are encouraged to contact the BTSC for more information on these organizations and for successful examples of Triad applications. More detailed information on DMA and on the Triad approach can be found in the Brownfields Technology Primer Series document *Using the Triad Approach to Streamline Brownfields Site Assessment and Cleanup*, which is available at www.brownfieldstsc.org. Project profiles, case studies, and other information on applying the Triad approach can be found at www.triadcentral.org. The BTSC provides other technical bulletins related to best practices embodied in the Triad approach such as "Use of Dynamic Work Strategies Under a Triad Approach for Site Assessment and Cleanup — Technology Bulletin." Additional documents providing critical information on related issues such as Green Remediation and Vapor Intrusion are also available.

REFERENCES

Brownfields and Land Revitalization Technology Support Center Web Site. 2008. Accessed August. www.brownfieldstsc.org

Fort Lewis, Washington. 2003. "Fort Lewis Agreed Order RI Demonstration of Method of Applicability Sampling and Analysis Plan Addendum Former Small Arms Ranges." October. Available at: www.triadcentral.org/user/doc/TPP-FortLewis-DMAMemo.pdf

Gilbert, R. 1987. "Statistical Methods for Environmental Pollution Monitoring. Wiley.

Lesnik, B. 2000. "Method Validation for the Resource Conservation and Recovery Act Program." *LC/GC Magazine*, Volume 18, Number 10. October.

Pacific Northwest National Laboratory. "Visual Sampling Plan Web page." Accessed August. Available at: http://vsp.pnl.gov

Triad Resource Center Web Site. 2008. Accessed July. Available at: www.triadcentral.org/index.cfm

U.S. Environmental Protection Agency (EPA). 2000. "Guidance for Data Quality Assessment: Practical Methods for Data Analysis" (QA/G-9). EPA 600-R-96-084. July. Available at www.epa.gov/oust/cat/epaqag9.pdf

EPA. 2003. "Using the Triad Approach to Streamline Brownfields Site Assessment and Cleanup." Brownfields Technology Primer Series. EPA 542-B-03-002. June. Available at: http://www.brownfieldstsc.org/pdfs/Triadprimer.pdf

EPA. 2004a. "Case Study of the Triad Approach: Expedited Characterization of Petroleum Constituents and PCBs Using Test Kits and a Mobile Chromatography Laboratory at the Former Cos Cob Power Plant Site." EPA 542-R-04-008.

EPA. 2004b. "Innovations in Site Characterization Case Study: Expedited Characterization and Excavation of Lead in Soil Using X-Ray Fluorescence at the Ross Metals Site, Rossville, Fayette County, Tennessee." December.

EPA. 2005a. "Understanding Dynamic Work Strategies Under a Triad Approach for Site Assessment and Cleanup—Technology Bulletin." September. Available at: www.brownfieldstsc.org/pdfs/DWSBulletin.pdf

EPA. 2005b. "Understanding Procurement for Sampling and Analytical Services Under a Triad Approach." EPA 542-R-05-022. June. Available at: www.brownfieldstsc.org/pdfs/procurement.pdf

EPA. 2005c. "Use of Field Portable X-Ray Fluorescence (FPXRF) and the Triad Approach to Investigate the Extent of Lead Contamination at a Small Arms Training Range, Fort Lewis, WA." August.

EPA. 2006a. "Data Quality Assessment: A Reviewer's Guide." EPA 240-B-06-002. February. Available at: www.epa.gov/quality/qs-docs/g9r-final.pdf

EPA. 2006b. "Data Quality Assessment: Statistical Methods for Practitioners." EPA 240-B-06-003. February. Available at: www.epa.gov/quality/qs-docs/g9s-final.pdf

EPA. 2006c. "Guidance on Systematic Planning Using the Data Quality Objectives Process." EPA/240/B-06/001. February. Available at: www.epa.gov/quality/qs-docs/g4-final.pdf

EPA. 2006d. "SW-846 Manual." Last updated October 18, 2006. Available at: www.epa.gov/epaoswer/hazwaste/test/sw846.htm

EPA. 2008a. "Environmental Technology Verification (ETV) Web page." Accessed August. Available at: www.epa.gov/etv

EPA. 2008b. "Performance-Based Measurement System (PBMS) Web page." Accessed August. Available at: www.epa.gov/SW-846/pbms.htm

EPA. 2008c. Information Quality Guidelines Web Site. Accessed July. Available at: www.epa.gov/quality/informationguidelines

NOTICE AND DISCLAIMER

This bulletin was prepared by EPA's Office of Solid Waste and Emergency Response under EPA Contract Nos. 68-W-02-034 and EP-W-07-078. The information in this bulletin is not intended to revise or update EPA policy or guidance on how to investigate or cleanup Brownfields or other revitalization sites. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This bulletin can be downloaded from EPA's Brownfields and Land Revitalization Technology Support Center at www.brownfieldstsc.org. For technical inquiries regarding this bulletin please contact Stephen Dyment of EPA at (703) 603-9903, or dyment.stephen@epa.gov.

Appendix 4

Calculation of Total Dioxin TEQs with Nondetect and Rejected Congeners

Helsel's Kaplan-Meier Approach

Calculation of sums or totals for multi-constituent chemicals (e.g., total dioxin Toxicity Equivalents (TEQs), total polychlorinated biphenyls (PCBs), total polynuclear aromatic hydrocarbons (PAHs), etc.) has typically involved simple substitution of zero, one-half the detection limit (DL), or the DL for left-censored (nondetect or less-than values) congeners. Because this practice introduces bias to estimates used in statistical calculations, however, many sources now strongly recommend against the use of arbitrary surrogate values for nondetects (Helsel 1990, 2005a, 2005b, 2009; EPA 2006, 2009b, 2009c).

Helsel (2009) describes an approach for calculating totals using the Kaplan-Meier (KM) product limit estimator, which is based on the following relationship between the "mean" of the toxic equivalence concentrations (TECs) and total TEQ for samples containing multiple congeners:

total concentration = "mean" TEC x n (where n is the number of congeners)

Note that this "mean" TEC is an intermediate value in the calculation that has no relationship to a mean TEQ for replicate decision unit (DU) samples. The KM estimator is a nonparametric maximum likelihood estimator that has been widely used in survival and failure analysis for more than 50 years (Kaplan and Meier 1958, Klein and Moeschberger 2003, Meeker and Escobar 1998). The KM estimator has only recently come into use in environmental assessment studies (Helsel 2005a), and is currently a default method used in EPA's ProUCL software for calculating the 95% UCL of the mean for data with one or more censored results (USEPA 2009b, 2009c).

Treatment of Nondetected Congeners

For the purposes of the UFP-QAPP template for dioxin soils assessment, the intermediate KM "mean" is recommended for use in calculating total dioxin TEQs, using the general equation presented above, in all cases where a) some fraction of the congeners are nondetect, and b) there are at least three detected congeners. Additional guidelines for calculating the KM intermediate "mean" are provided below. If all congeners are detected, then the intermediate "mean" calculated by the equation is the arithmetic average of all the congeners' TECs.

If only one or two congeners are detected, then there is no statistically satisfactory method for calculating the dioxin TEQ that adequately accounts for the uncertainty introduced by nondetect congener results. In this case, the intermediate "mean" should be calculated as the arithmetical average, where simple substitution is used for nondetects. A quasi-sensitivity analysis approach is recommended, wherein substitution of both zero and the DL are used to calculate lower- and upper-bound estimates for the total TEQ. Compare the TEQs from both approaches to assess whether they have the same decision outcome. Substitution of one-half the DL can be used to calculate a "middle-

of-the-road" value, although it should be acknowledged that the uncertainty of this estimate may be unacceptable for decision making.

In cases where critical decisions hinge on total TEQ estimates with mostly nondetect results, project teams are advised to consider:

- consulting personnel with expertise in statistics,
- reanalyzing existing samples (if archived samples are available and meet holding times),
- comparing with results from nearby similar DUs and the CSM, or
- collecting additional samples.

The stepwise KM approach for calculating the total dioxin TEQ for individual samples is described below:

Step 1. Calculate the TEC for each congener by multiplying the results for individual congeners by their congener-specific TEF (van den Berg and others 2006). For nondetect congeners, the reporting limit or DL should be multiplied by the TEF.

Step 2. Calculate the intermediate "mean" TEC for each sample using a KM calculator spreadsheet. If all the congeners are detected, then calculate the intermediate value as the arithmetic mean. If nondetects are present and at least three results are detected, calculate the KM intermediate using one of the options described below. If only one or two congeners are detected, use simple substitution and a quasi-sensitivity analysis approach, as discussed above.

Step 3. Calculate the total dioxin TEQ using: Total TEQ = intermediate "mean" TEC x n, where n is the number of congeners in the calculation.

Helsel (2009) discusses several potential contraindications for calculation of the KM mean. The first concerns cases where only a single DL is used for all nondetect congeners. This is not expected to occur for calculation of total dioxin TEQs, since results for individual congers are first multiplied by congener-specific TEFs. The second contraindication is when the maximum reported result is a nondetect, high-toxicity (i.e., TEF close to 1) congener. This is problematic, as the KM method effectively ignores maximum results that are censored. Helsel (2009) suggests that the DL be substituted in these cases, but that it should be acknowledged that this represents a worst-case scenario. Another option is to compare the congener concentration and congener profile of the sample with a high TEF nondetect to results from similar (per the CSM) DUs. If the congener profiles are similar, but the other DUs have a detection for the congener in question, substitution of a value (straight substitution, an average of several, or a maximum) from the other DUs may be made.

Helsel (2009) does not discuss the minimum number of detected results required to estimate the KM mean, but a practical minimum of three detected results is recommended. Cases where only one or two congeners are detected are discussed above. Lastly, Helsel (2009) recommends that for left-censored environmental data, Efron's bias correction should always be used. This simply requires that the minimum result always be treated as a detected result. The manner in which

Efron's bias correction is incorporated in calculations of the KM mean depends on the specific software or approach used. For example, for programs that require a "flag" to distinguish between detected and nondetect data, one only needs to use the appropriate flag for detected data to qualify the minimum result(s).

Three options are described below for calculation of the KM mean:

(1) Helsel's KM Excel spreadsheet model (available from www.practicalstats.com). This spreadsheet has been built into a workbook designed specifically for calculating the TEQ from raw data congener concentration data. Raw data are entered into one spreadsheet, which automatically calculates the toxic equivalent concentration (TEC) for each congener. The TECs are copied and pasted into a second spreadsheet in the workbook that performs the KM calculation. This produces an intermediate value (the KM "mean") which is transferred back to the first spreadsheet. The intermediate result is then automatically multiplied by the number of congeners to produce the total TEQ for the sample. Detailed instructions for using the spreadsheets are included in the Excel workbook's spreadsheets.

(2) Alternatively, EPA's ProUCL software may be used. Before estimates of the KM intermediate "mean" TEC can be calculated, the congener concentration results (in parts per trillion (ppt)) are converted to congener TECs by multiplying each congener by its TEF. This should be done independently before the TECs are put into ProUCL for the KM calculation. (ProUCL cannot do the TEC calculation.) The TECs are then entered into ProUCL and the KM intermediate "mean" is automatically calculated for data sets with one or more nondetect results. EPA (2009b, 2009c) should be consulted for instructions when entering data into ProUCL, since a coding procedure is used in ProUCL to "tell it" which congener TECs were from ND values. Note that in order to use Efron's bias correction, the minimum result should be coded as a detected result. If intermediate "means" are required for multiple samples, then each sample needs to be identified using a "grouping" variable (see EPA 2009b). For each sample, the KM intermediate "mean" needs to be extracted from the ProUCL report, and manually multiplied by the number of congeners to produce the total TEQ result for that sample.

(3) Commercial or other statistical software. The KM model is included in many mainstream statistical software packages, as well as public domain (including the R language) programs. Helsel (2005a) discusses an approach for "flipping" data for use in commercial packages, which emphasize treatment of right-censored data. Experienced users may elect to use alternative approaches for calculation of the KM intermediate "mean," but should use methods employing Efron's bias correction, and demonstrate that results are comparable to the intermediate "means" calculated using Options (1) or (2) above.

Treatment of R-Qualified Congeners

One additional component for assessing the uncertainty of estimates of the intermediate KM "mean" and total TEQ, concerns treatment of rejected (R qualified) data. It is possible to reject individual congener analytes based on ion abundance, the signal-to-noise ratio, relative retention time, a low laboratory control sample result, gross blank contamination, or other analyte-specific criteria. For individual chemicals with replicate samples (i.e., sufficient sample-sizes to support

calculations), rejected data are always excluded from calculations in environmental assessments. However, for calculation of the "mean" (and total) for a set of congeners, there is concern that exclusion of rejected data may bias estimates low or create a need for replacement data (resampling or reanalysis). The magnitude (and importance) of this bias depends on the values reported for R-qualified data, as well as the congener-specific TEFs.

Although rejected data should not be included in final calculations of TEQ for a given sampling or decision unit, rejected data values (concentrations or detection limits) can be included in KM "mean" and total TEQ calculations early in the data evaluation process. These TEQs can be compared to TEQs calculated with the rejected values removed. This quasi-sensitivity approach, similar to that recommended above for nondetect values, can assist project teams in assessing the magnitude of impacts from rejected data and the need for replacement data (Replacement data may require reanalysis of samples at the laboratory, with laboratory corrective actions or method refinements as needed, or the collection of additional samples from the site). Rejected data can be further evaluated through professional judgment, such as whether a rejected congener may be present at a concentration that could affect the TEQ based on historical site information or data from surrounding decision units. For example, project teams could use the KM calculator to further assess how high the concentration of a rejected congener would have to be to affect the TEQ, and then compare this estimate to concentrations for this congener that are present in other decision units, or in comparable historical data sets.

Treatment of EMPC values and qualified data

The Contract Laboratory Program Statement of Work (CLP SOW) for dioxin analysis specifies the reporting of detected congeners as "EMPC" values ("estimated maximum possible concentration") when a congener peak is present at an acceptable signal-to-noise ratio, but ion abundance criteria are not met for definitive identification of that congener. The CLP SOW excludes these values from the calculation of TEQ. EPA Method 8290A also specifies the reporting of EMPC values but makes no recommendations concerning their use in TEQ calculations. EMPC values are generally qualified as estimated concentrations ("J") or nondetect values ("U") during data validation in accordance with EPA Functional Guidelines. When qualified "J", EMPC values can be applied along with other J-qualified congener results in TEQ calculation and risk assessment (J-qualified data are generally applied like unqualified data under EPA risk assessment protocols). EMPC values qualified "U" can be treated as other nondetect values using the KM approach described above. Given that use of EMPC values may overestimate the TEQ and associated dioxin risk, project teams may again elect to perform a quasi-sensitivity analysis by calculating TEQ without the EMPC values. As for rejected data, significant effects from EMPC values may require corrective action to improve data quality (such as sample reanalysis).

Therefore, for congeners that are influential (high-toxicity, TEF close to 1, or high concentration) in calculations of the intermediate "mean" and total TEQ, rejected and qualified data may require further evaluation by project teams. The uncertainty of calculating total TEQs, as can be demonstrated through sensitivity analyses, should be addressed in the uncertainty section of assessment documents, and taken into account in decision-making.

Appendix 5

References

Arjomand, L.H. 1997. On-line Statistics Course (BUSA 3101), Lecture Notes 7, Sampling Distribution of the Mean. Clayton State University School of Business. Last update on 7/23/2010. http://business.clayton.edu/arjomand/business/l7.html

Hathaway, J.E., G.B. Schaalje, R.O. Gilbert, B.A. Pulsipher, and B.D. Matzke. 2008. Determining the Optimal Number of Increments in Composite Sampling. Environ. Ecol. Stat. 15:313–327.

Hawaii Department of Health (DOH); 2009. www.hawaiidoh.org/tgm.aspx?p=0402a.aspx

Helsel, D.R. 1990. Less than obvious, statistical treatment of data below the detection limit. Environmental Science and Technology. Volume 24, Number 12. Pages 1767 through 1774.

Helsel, D.R. 2005a. Nondetects and Data Analysis. Statistics for Censored Environmental Data. John Wiley and Sons, Inc. Hoboken, NJ. 250 pages.

Helsel, D.R. 2005b. More than Obvious: Better Methods for Interpreting Nondetect Data.. Environmental Science and Technology. Volume 39, Number 20. Pages 419A through 423A.

Helsel, D.R. 2009. Summing Nondetects: Incorporating Low-Level Contaminants in Risk Assessment. Integrated Environmental Assessment and Management. Volume 6, Number 3. Pages 361 through 366.

Hewitt, Alan D,. T.F. Jenkins, M. Walsh and S. Brochu. 2009. Environmental Security Technology Certification Program (ESTCP), Validation of Sampling Protocol and the Promulgation of Method Modifications for the Characterization of Energetic Residues on Military Testing and Training Ranges. ESTCP Final Report ER-0628. June. http://www.estcp.org/viewfile.cfm?Doc=ER%2D0628%2DFR%2Epdf

ITRC. 2003. Technical and Regulatory Guidance for the Triad Approach: A New Paradigm for Environmental Project Management. Interstate Technology and Regulatory Council. December. http://www.itrcweb.org/Documents/SCM-1.pdf

Jenkins, T., et al. 2005. Representative Sampling for Energetic Compounds at Military Training Range: Environmental Forensics, Volume 6, pp. 45-55. http://www.environmentalforensics.org/journal.htm.

Kaplan, E.L., and P. Meier. 1958. Nonparametric Estimator from Incomplete Observations. Journal of the American Statistical Association. Volume 53. Pages 457 through 481.

Klein, J.P., and M.L. Moeschberger. 2003. Survival Analysis – Techniques for Censored and Truncated Data. 2nd Edition. Statistics for Biology and Health, Springer Verlag. 560 pages.

Matzke, B.D., Nuffer, L.L., Gilbert, R.O., Hathaway, J.E., Wilson, J.E., Dowson, S.T., Hassig, N.L., Murray, C.J., Pulsipher, B.A., and S. McKenna. 2007. Visual Sampling Plan Version 5.0 User's Guide. PNNL-16939. Pacific Northwest National Laboratory, Richland, WA. September.

Meeker, W.Q., and L.A. Escobar. 1998. Statistical Methods for Reliability Data. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. New York, New York. 712 pages.

Ramsey, C. and A. Hewitt. (2005). A Methodology for Assessing Sample Representativeness, Environmental Forensics 6:71-75. http://www.environmentalforensics.org/journal/volume6/number2.htm.

Rohlf, F. J., H. R. Akcakaya, and S.P. Ferraro. 1996. Optimizing composite sampling protocols. Environmental Science and Technololgy 30(10): 2899-2905.

U.S. Environmental Protection Agency (USEPA). 1990. A Rationale for the Assessment of Errors in the Sampling of Soils. EPA/600/4-90/013. Environmental Monitoring Systems Laboratory. Las Vegas, Nevada. May.

United Nations Environment Programme (UNEP). 2007. World Health Organization re-evaluation of dioxin toxic equivalency factors. UNEP/POPS/COP.3/INF/27. April. http://www.pops.int/documents/meetings/cop_3/meetingdocs/inf27/K0761219%20POPS-COP3-INF27.pdf

USEPA. 1989a. Soil Sampling Quality Assurance Guide. EPA/600/8-89/046. March. http://www.epa.gov/esd/cmb/research/bs122.pdf

USEPA. 1989b. Methods for the Attainment of Cleanup Standards, Vol.1 Soils & Solid Media. EPA 230/02-89-042. February. http://www.epa.gov/tio/download/stats/vol1soils.pdf

USEPA. 1989c. Risk Assessment Guidance for Superfund, Vol. 1, Part A. EPA 540/1-89/002 December. http://epa.gov/oswer/riskassessment/ragsa/index.htm

USEPA. 1990. A Rationale for the Assessment of Errors in Soil. EPA 600/4-90/013. May. http://clu-in.org/download/stats/rationale.pdf

USEPA. 1991. Guidance for Data Useability in Risk Assessment (Part A). EPA/540/R-92/003. December. http://www.epa.gov/oswer/riskassessment/datause/parta.htm

USEPA. 1992a. Supplemental Guidance to RAGS; Calculating the Concentration Term. OSWER 9285.7-081. May. http://www.deq.state.or.us/lq/pubs/forms/tanks/UCLsEPASupGuidance.pdf

USEPA. 1992b. Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies. EPA/600/R-92/128. Office of Research and Development. Washington, DC. July.

USEPA. 1993. Guidance for Evaluation of the Technical Impracticability of Ground-water Restoration. EPA/540-R-93-081. September. http://www.epa.gov/superfund/resources/gwdocs/techimp.htm

USEPA. 1994. Statistical methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media. EPA 230-R-94-004. U.S. Environmental Protection Agency, Office of Policy, Planning and Evaluation. Washington, DC. June.

USEPA. 1995. Superfund Program Representative Sampling Guidance, Vol. 1: Soil (Interim Final). EPA/R-95/141. December. http://cluin.org/download/char/sf_rep_samp_guid_soil.pdf

USEPA. 1996a. Soil Screening Guidance: User's Guide. Superfund Publication 9355.4-23. EPA/540/R-96/018. July. http://www.epa.gov/superfund/health/conmedia/soil/index.htm

USEPA. 1996b. Soil Screening Guidance: Technical Background Document. EPA 540/R-95/128. May. http://www.epa.gov/superfund/health/conmedia/soil/index.htm

USEPA. 1996c. PCBs: Cancer Dose-Response Assessment and Application to Environmental Mixtures. EPA-600/P-96/001F. September. http://www.epa.gov/osw/hazard/tsd/pcbs/pubs/pcb.pdf

USEPA. 1999. Correct Sampling Using the Theories of Pierre Gy. Office of Research and Development Technology Support Project. March. http://www.epa.gov/esd/factsheets/csutpg.pdf

USEPA. 2000. EPA Quality Manual for Environmental Programs. CIO 2105-P-01-0. May. http://www.epa.gov/irmpoli8/policies/2105P010.pdf

USEPA. 2001a. Communicating Risks and Uncertainties in Probabilistic Risk Assessments. RAGS Vol.3 Part A. Ch 6. www.epa.gov/riskassessment/rags3adt/pdf/chapter6.pdf

USEPA. 2001b. Risk Assessment Guidance for Superfund, Part D. Publication 9285.7-47. December. http://epa.gov/oswer/riskassessment/ragsd/index.htm

USEPA. 2002a. Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites. EPA 540-R-01-003, Office of Emergency and Remedial Response, Washington, DC. September.

USEPA. 2002b. Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan, EPA QA/G-5S. EPA/240/R-02/005. December. http://www.epa.gov/quality/qs-docs/g5s-final.pdf

USEPA. 2002c. Supplemental Guidance for Developing Soil Screening Levels for Superfund Sites. OSWER 9355.4-24. December. http://www.epa.gov/superfund/health/conmedia/soil/index.htm

USEPA. 2002d. Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites. EPA 540-R-01-003. OSWER 9285.7-41. September. http://www.epa.gov/oswer/riskassessment/pdf/background.pdf

USEPA. 2003. Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples. EPA 600/R-03/027. November. http://www.epa.gov/tio/download/char/epa_subsampling_guidance.pdf

USEPA. 2006. Guidance on Systematic Planning Using the Data Quality Objectives Process, EPA QA/G4, EPA/240/B-06/001, Office of Environmental Information, Washington, DC. February. http://www.epa.gov/quality/qs-docs/g4-final.pdf

USEPA. 2006. On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations. Singh, A., Maichle, R., and S.E. Lee. EPA/600/R-06/022. March.

USEPA. 2009a. ProUCL Version 4.00.02 User Guide (Draft). Singh, A., R. Maichle, A.K. Singh, S.E. Lee, and N. Armbya. Office of Research and Development, National Exposure Research Laboratory, EPA/600/R-07/038. February.

USEPA. 2009b. ProUCL Version 4.00.04 Technical Guide (Draft). Prepared by A. Singh and A.K. Singh. EPA/600/R-07/041. February.

USEPA. 2009c. ProUCL Version 4.00.04 User Guide (Draft). Singh, A., R. Maichle, A.K. Singh, S.E. Lee, and N. Armbya. Office of Research and Development, National Exposure Research Laboratory. EPA/600/R-07/038. February.

USEPA. 2009d. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities. Unified Guidance. EPA 530-R-09-007. Office of Resource Conservation and Recovery. March.

Van den Berg, M. and others. 1998. Toxic Equivalency Factors (TEFs) for PCBs, PCDDs, PCDFs for Humans and Wildlife. Environmental Health Perspectives. Volume 106, Number 12. December.

Van den Berg, M. and others. 2006. The 2005 World Health Organization Reevaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-Like Compounds. Toxicological Sciences. Volume 93, Number 2. Pages 223 through 241. On-Line Address: http://epa-prgs.ornl.gov/chemicals/help/documents/vandenberg2006.pdf