EPA NEW ENGLAND

QUALITY ASSURANCE PROJECT PLAN GUIDANCE FOR

ENVIRONMENTAL PROJECTS USING ONLY

EXISTING (Secondary) DATA

U.S. EPA NEW ENGLAND

Quality Assurance Unit
Office of Environmental Measurement and Evaluation

Revision 2

Quality Assurance Project Plan Guidance
For Environmental Projects Using Only Existing (Secondary) Data

Purpose

EPA policy requires that all environmental data used by the Agency must be known and documented quality. This includes new data generated for a project and existing data that were previously collected for other projects. The purpose of this Guidance is to describe project information that should be included in a quality assurance plan for environmental projects based solely on existing data. Specifically, it describes a streamlined approach for planning and documenting projects that do not include sampling, testing or measurement determinations.

Using Existing (Secondary) Data in Environmental Projects

Environmental data projects typically involve planning, sampling, analysis, assessment and data review. In planning their investigations, project teams generally use existing data to develop sampling designs and to decide how much and what type of data to collect. The term existing data is used interchangeably with "secondary data" and "non-direct measurements." Existing data may come from a number of sources, including other studies, government databases, etc. The original purpose for collecting these secondary data may be very different from that of the current investigation. Also, these secondary data may have been collected using different sampling methods (composite vs. grab, random vs. hot spot sampling), and/or analytical methods than those selected for the current project.

Basing project decisions on existing data may result in errors if secondary data were not generated for the same purpose or using the same methods as the current investigation. Data could be biased and final conclusions could be impacted.

Therefore, before using secondary data, project team members should evaluate the data to identify any limitations on their use. Also, to ensure transparency in decision making, criteria and reasons for ***including*** and ***excluding*** certain data from use must be clearly documented. Failure to clearly document why data are included or excluded can result in the appearance of biased data selection and diminish the product's credibility.

Project personnel should describe the processes for selecting and for evaluating existing data in the quality assurance plan in accordance with ***EPA Requirements for Quality Assurance Project Plans*** QA/R-5 http://www.epa.gov/quality/qs-docs/r5-final.pdf .

For an in-depth discussion on when and how to use existing data in environmental projects, refer to EPA Guidance for Quality Assurance Project Plans QA/G-5 "Chapter 3: Projects Using Existing Data" http://www.epa.gov/quality/qs-docs/g5-final.pdf

Sources of secondary data include the following:

- Environmental indicator data obtained from federal/state/local databases and records
- Existing sampling and analytical data from a previous investigation of the area
- Computer model simulations and applications pertaining to other studies
- Historical data (e.g., from organization's/facility's corporate records and/or federal/state local records pertaining to previous monitoring events, site assessments, investigations, etc.)
- Background information/data from organization's/facility's corporate records and/or federal/state/local records pertaining to site-specific industrial processes, process by-products, past and current chemical uses, raw material and finished product testing, waste testing and disposal practices, and potential chemical breakdown products
- Data generated to verify innovative technologies and methods
- Data obtained from computer databases (such as manufacturers' process/product information, waste management or effluent information, and EPA or state data bases)
- Literature files/searches
- Publications
- Photographs
- Topographical maps
- Meteorological data

Projects Based Only on Existing Data

Although most environmental projects involve the generation of new data, projects based solely on the use of existing data are increasingly common. Re-purposing existing data saves time, resources and may resolve sampling access problems.

For projects using only existing data, EPA New England requires that ***Secondary Data Quality Assurance Project Plans*** **(QAPPs)** be developed and submitted to EPA for review and approval prior to the start of the project. For these projects, a QAPP should be prepared using a streamlined (graded) approach. For example, certain QAPP elements specific to sampling would not be applicable; therefore, certain standard QAPP sections could be omitted.

Note: If secondary data will be used in the development, evaluation and/or application of

environmental models, then the project team should follow the template and checklist provided at the Region 1 web site http://epa.gov/ne/lab/qa/qamodeling.html

A graded approach to ***Secondary Data QAPPs*** would include the following QAPP elements:

1.0 PROJECT MANAGEMENT – ORGANIZATION AND RESPONSIBILITIES

1.1 **Title and Approval Page** (including signature dates). Include the following:

- EPA Project Officer and EPA QA Officer
- Project Officer and quality personnel from funded organization
- Other responsible project team members including contractors, consultants, voluntary organizations, etc.

1.2 **Table of Contents**

1.3 **QAPP Distribution List**. Include all personnel accountable for the outcome of the project; involved in gathering and evaluating secondary data; and project personnel who will ultimately use the project results. Include contact information.

1.4 **Project Organization.** Identify key project team members and their organizations. Include those responsible for selecting, compiling and evaluating existing data. Also identify those responsible for project planning, coordination, data analysis, report preparation, and quality assurance. Provide an organizational chart showing lines of communication.

Describe any specialized training or qualifications needed by team members to obtain and analyze existing data.

1.5 **Purpose of Study, Background Information, and Problem Definition.** Clearly state the reason for conducting the project. Discuss the desired outcome of the project in terms of decisions that can be made or actions that can be taken. Provide enough background information to put the project in programmatic context and to explain the environmental problem.

If this information is provided in another document (grant proposal, scope of work, etc.), reference the document and provide it as an attachment to the QAPP.

1.6 **Overview of Project Tasks.** Describe planned data activities including how existing data will be used to investigate the current environmental problem.

- Describe type and amount of data that will gathered (e.g., age of data, geographical representation, temporal representation)

- Explain how data sources will be selected or rejected for use
- Describe approach for analyzing data including formulas, calculations, units, definitions of terms, and statistical analysis, will be included and defined.

Similarly, if this information is provided in another document (grant proposal, scope of work, etc.), reference the document and provide it as an attachment to the QAPP.

1.7 **Quality Objectives and Criteria:** State the overarching quality objectives that must be met to ensure a successful outcome of the project. The quality objectives of the project are determined by the end users (e.g., risk assessors, regulators, local state government, citizen groups, etc.). For example, if the project quality objective is to compile and analyze scientifically sound, defensible and transparent data adequate for the development of a TMDL, then acceptance criteria used for evaluating the quality of existing data must be tight enough to minimize decision errors.

Specify acceptance criteria for each matrix and measurement (analytical) parameter and indicate QC sample or activity associated with the quality indicator. For example:

Matrix: Stream water
Measurement Parameter: Nitrate – Nitrogen

- ✓ Precision (e.g., Relative Percent Difference $\leq$20%, field duplicates)
- ✓ Precision (e.g., Relative Percent Difference $\leq$15%, laboratory duplicates)
- ✓ Accuracy (e.g., 85 – 115%, Spiked Control Samples)
- ✓ Sensitivity (0.05 mg/L, calibration standards)
- ✓ Comparability (all nitrate analyses generated in accordance with USEPA Method 300.1, Method citation)
- ✓ Representative sampling (Documented sampling SOPs must used by trained personnel, required training documentation)

Quality acceptance criteria are unique to each project; many environmental studies (e.g., non-TMDL projects) may require less stringent data quality acceptance criteria.

Once data acceptance criteria are established, the project team selects existing data that meet the criteria. In order to determine the quality "pedigree" and usefulness of the secondary data, supporting QC information must be reviewed. Information (a.k.a., metadata) about why, how, and when the existing data were collected provides the user with more confidence. Metadata are documented in project reports, validation reports and accompany database information.

2.0 DATA SELECTION AND MANAGEMENT

2.1 **Sources of Existing Data.** List the sources(s) of all secondary data that may be used, including:

- type of data and collection dates
- originating organization
- report title, author and date
- data base names

Note: Information may be presented in tabular format; an example is provided in Attachment 1.

Explain the reason(s) for selecting various sources(s) of existing data (data bases, reports, etc.).

To ensure transparency and defensibility in the decision making process, it is very important to document why certain ***related*** project reports and/or existing data were ***not*** used. For example, if reported dry weather data did not have associated information on number of antecedent dry days prior to collection, then the project team may decide not to use the data.

Describe the data format (e.g., electronic, hardcopy) and how data will be maintained for the project. If data are obtained from data bases, include as much accompanying quality control, temporal, locational data, etc. as needed to document and verify the quality of the data.

2.2 **Intended Use of Existing Data.** Describe how different types of data will be used. For example, certain data may be used to define the boundaries of a contaminated area, while other data may be used in identifying other suspected pollutants or breakdown products.

State how and when data that are found to have limitations (e.g., lab qualified data) will be used in the project.

2.3 **Limitations on the Use of Existing Data.** Specify criteria for selecting existing data for the project. Appropriate selection criteria will ensure that secondary data are "good enough" to support project conclusions, decisions or actions. Selection criteria support the general project quality objectives described in Section 1.7.

The following are some examples of selection criteria. **Note: These are examples; and may not apply to your project.**

- Data sets must include quality control (QC) metadata for precision and accuracy.

- Data must be generated under an approved QAPP or other sampling document
- Analytical methods must be sufficiently sensitive (i.e., low enough reporting limits) to support data reporting to state water quality criteria levels.
- All existing data sets used in the project must be generated using the same or comparable sampling and analytical methods or SOPs.
- Data must indicate if results are from composite or grab sampling.
- Sampling design must identify samples that were collected using statistical approach, i.e., "hot spot", random, or grid.
- Reported data must include laboratory qualifiers and qualifier definitions
- Dry weather data used in the project must be preceded by a minimum 72-hour dry period.
- Only data generated after 1/1/2000 will be used for the current investigation.

Identify personnel responsible for selecting project data and the process used.

If no known quality requirements were applied during the sampling and analysis of the existing data, then state this in the QAPP. A disclaimer should be added to any project deliverable to indicate that the quality of the secondary data is unknown. Include the wording for the disclaimer in the QAPP.

3.0 ASSESSMENTS AND OVERSIGHT

3.1 Explain how the project team will ensure that project tasks are completed as planned. Identify personnel responsible for conducting audits and/or overseeing the project.

3.2 Indicate how project oversight will be documented (e.g., assessment reports, memos, etc.)

3.3 Describe how problems will be resolved, including chain-of-command, and documentation process. Include examples of types of corrective actions that might be implemented (e.g., access other data sources, loosen or tighten acceptance criteria).

4.0 DATA REVIEW - VERIFICATION, VALIDATION AND EVALUATION

4.1 Describe how project members will review and verify or validate the adequacy of each data set data relative to the established acceptance criteria established in Section 1.7. Describe the following:

- how data qualifiers will be applied to data not meeting project acceptance criteria. Define laboratory and validation qualifiers (e.g., U, B, J, R, etc.)
- when data will be rejected (not used), and
- how limitations on the use of individual data sets will be documented.
  Note: Typically this is done in the Quality Control Section of the Final Report or

Project Deliverable.

4.2 Finally, describe how all the gathered data will be evaluated to ensure they can be used for project purposes. Describe any statistical applications used to identify outliers, etc.
Consider the following issues when reconciling data with the project objectives.

- are the data unbiased and sufficiently representative to be used for the current project?
- are data sets complete?
- were data sets collected using the same or comparable methods or SOPs?
- were data collected and managed according to an approved QAPP?
- were data collected by trained personnel familiar with the appropriate SOPs?
- do data meet necessary detection limits and are they reported in the right units of measurement?

5.0 PROJECT SCHEDULE

Include a project time line/schedule including dates for meetings, product deliverables, and final report.

6.0 PROJECT REPORTING

Describe how project results will be reported (e.g., report, deliverable document, etc.). Provide the proposed document outline that includes a quality assurance section.

Also include a statement in this Section that the final project report will identify all sources of existing data that were used in the project, and that they will be either provided as attachments to the Final Report, available through embedded websites, or available upon request.

REFERENCES

1) ***QAPP Requirements for Secondary Data Research Projects***
Example Guidance provided by EPA National Risk Management Research Laboratory, July 1999 http://www.epa.gov/quality/qs-docs/found-data-qapp-rqts.pdf

2) ***EPA Requirements for Quality Assurance Project Plans*** QA/R-5
http://www.epa.gov/quality/qs-docs/r5-final.pdf

3) ***EPA Guidance for Quality Assurance Project Plans*** QA/G-5 "Chapter 3: Projects Using Existing Data" http://www.epa.gov/quality/qs-docs/g5-final.pdf

4) Workbook for ***Uniform Federal Policy for Quality Assurance Project Plans***, Version 1 March 2005, http://www.epa.gov/fedfac/pdf/Wkbk_Mar05.doc

Attachment

Example UFP QAPP Worksheet #13 (*Uniform Federal Policy for Quality Assurance Plans*, Version 1 March 2005, http://www.epa.gov/fedfac/pdf/Wkbk_Mar05.doc

Identify information and/or data generated/collected outside of the current project activity that will be used to make environmental decisions for the project. Specify how those acquired data/information will be used and the limitations on their use. These limitations include data quality considerations/ problems as well as documentation completeness.

Non-Direct Measurement (Secondary Data)	Data Source (Originating Organization, Report Title and Date)	Data Generator(s) (Originating Org., Data Types, Data Generation/Collection Dates)	How Data Will Be Used	Limitations on Data Use
Soil Gas Data	*BioWatch Consulting, LTD: "Titanic Shipyard Investigation Report," 11/20/95*	*BioWatch Consulting, LTD: VOC Soil Gas Data, Sample Collection Dates: 10/19-23/95*	*To assess the potential sources of contaminated soil and resultant groundwater migration*	*1. Unvalidated data used to generate report 2. Insufficient data points to fully characterize on-site contamination and off-site migration*
Municipality Drinking Water Data	*XYZ Municipality: Quarterly Drinking Water Check Report, 6/95 - 6/96*	*Smith Laboratories, Inc.: VOC Drinking Water Data, Sample Collection Dates: 6/12/95, 9/15/95, 12/10/95, 3/6/96, 6/12/96*	*To assess existing groundwater contamination*	*1. Unvalidated data used to generate report 2. Limited number of wells exist to sample*