

Challenges of Planning and Integrating Data Reporting, Verification, and Storage, for a Large Scale Sediment Remediation Program

Louis Blume
USEPA Great Lakes National Program Office Chicago, IL

Judy Schofield, Neal Jannelle, Harry McCarty, and
Molly Amos
Computer Sciences Corporation
Alexandria, VA

Mark Kromis
EarthTech
Roswell, Georgia

April 23, 2008



Workshop Overview



Standardizing Data Reporting, Review and Storage: Examples from the Front Lines

- Lou Blume
 - US EPA Great Lakes National Program Office
- Michael Johnson
 - US EPA Analytical Services Branch
- Bob Runyon
 - US EPA Region 2
- Bruce Means
 - US EPA Analytical Services Branch

Presentation Overview



- Background on Great Lakes Legacy Act (GLLA)
- Challenges in managing data and meeting quality system goals under the GLLA
- Data Management procedures under the GLLA
- Lessons learned
- Recommendations

The data management challenges that GLNPO is facing are similar to those that the Agency will face in the event of a national emergency.

Great Lakes Legacy Act of 2002



...a new “tool” in
the Great Lakes
sediment
remediation toolbox

History



- "Great Lakes and Lake Champlain Act of 2002" passed by Congress on November 12, 2002
- Signed into law by President Bush on November 27, 2002 (Public Law No: 107-303)
- Authorizes \$50M per year from FY2004 through FY2008 for contaminated sediment projects in the Great Lakes

Appropriations to date



- FY2004 \$9.9 Million
- FY2005 \$22.3 Million
- FY2006 \$29.3 Million
- FY2007 \$30 Million
- FY2008 \$34.5 Million

Areas of Concern



- Currently 42 Areas of Concern (AOCs) throughout the Great Lakes
- Contaminated sediments can be found in many of these AOCs, resulting in various beneficial use impairments
- Since 1997, approximately 3.3 million cubic yards of contaminated sediments have been remediated in the Great Lakes AOCs

Cost-sharing



- Non-Federal cost share must be at least 35% of total project costs and 100% of operations and maintenance costs, and:
 - May include in-kind services,
 - May include monies and in-kind services under an administrative order on consent or judicial consent decree,
 - May NOT include any funds pursuant to a unilateral administrative order or court order.

GLLA Project Agreement

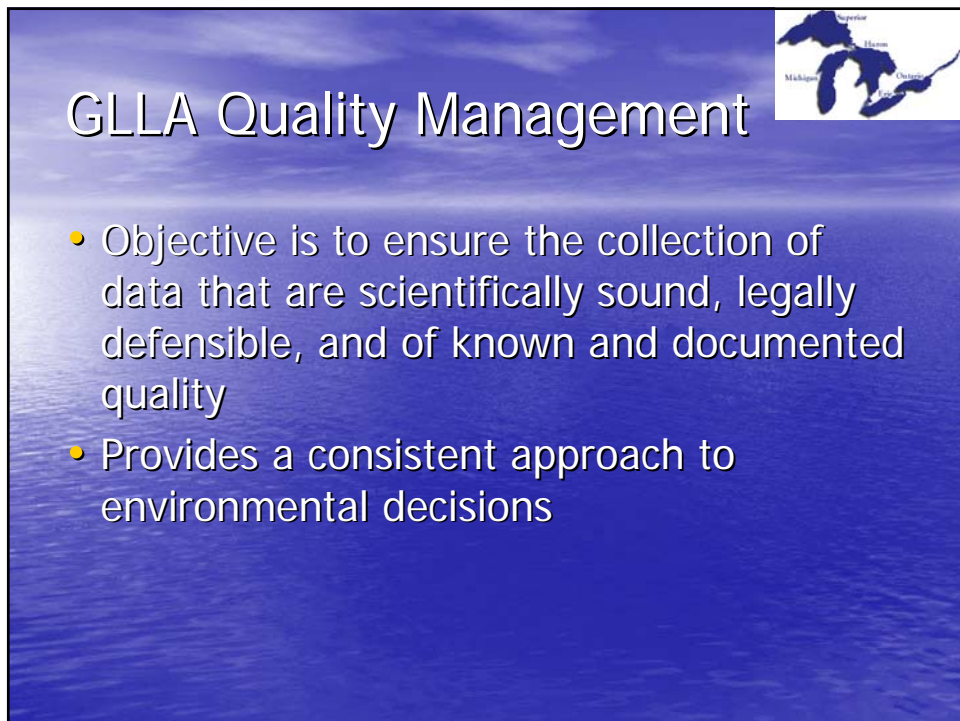


- Not a grant; not a contract
- Project Agreement (PA)
- A negotiated agreement between EPA and the non-Federal Sponsor
- Outlines roles and responsibilities of each party

GLLA Partners



- Non-Federal Sponsors
 - States
 - Potentially responsible parties (PRPs)
 - Local consortiums
- Contractors
 - Remedial Action Contracts (RACs)
 - Emergency and Rapid Response Services (ERRS)
 - Superfund Technical Assessment & Response Team (START)
- Cooperators
 - US Army Corps of Engineers
 - Department of Energy, Battelle
 - EPA Regions



GLLA Quality Management

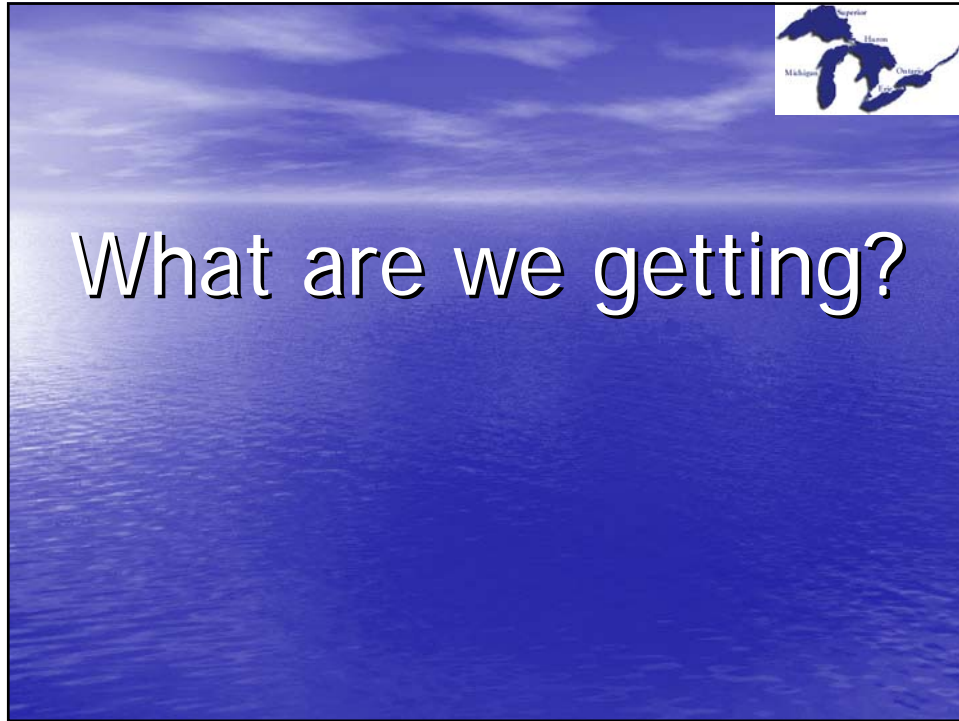


- Data Management must be “cradle to grave”
 - Project is not truly completed until project information is compiled and available to stakeholders and the public
 - Project information must be transparent and reproducible
 - Completion of project reports are driving the process

GLLA Data Management



- GLLA QMP stipulates that
 - sufficient documentation be provided with submission of a data set to assist data users when evaluating the utility of the data set for their purposes
 - includes the original information on data quality associated with the data
 - supplementary information including data verification narratives
 - quality documentation for each project addresses data management issues including collection, reporting, verification, storage, and usability



GLLA Data Management Challenges

- Initial data reporting is lacking quality control (QC) information
- Geographic coordinates of various sampling locations had limited QC
 - Field data submitted in numerous formats and often lacked reference points
 - When plotting sample coordinates observed numerous errors
- Lack of QC information for engineering aspects of projects such as volume estimates

GLLA Data Management Challenges



- Lack of statistical sampling design
 - Often stratified dots on a map
- Rationale of sampling design is not documented for historic monitoring
- Data usability assessment impossible without proper documentation of basis of original sampling design
- Lack data quality assessment in the context of project goals
 - Emphasis is on QC for an individual sample
 - Did data as a whole meet project requirements?

GLLA Quality Management Challenges



- The logistics of sediment remediation projects and the complex array of roles, responsibilities, and funding sources complicate the quality management process thus:
 - Participation of Quality Manager as a true member of the Project Team
 - Quality program is built on regular communication among stakeholders

GLLA Data Management Challenges



- Dealing with multiple entities with distinct lines of authority
- Must work within existing vehicles and contract requirements
- Final project decisions and products must meet Information Quality Guidelines (IQGs)
 - ensure and maximize the quality of information
 - provide a transparent process, and products that are clear, consistent, and reasonable

GLLA Data Management Challenges



- Existing site assessment data
 - Incomplete field information
 - Geographic referencing
 - Analytical Data of known and documented quality
 - Understanding level of verification often impossible
 - Sensitivity of methods
 - Reporting of non-detect data
 - Reproducibility of data
 - Method differences

GLLA Data Management Challenges



- Remedial projects
 - Evaluating achievement of clean up goals must happen real time
 - Requires quick turnaround data generation
 - Short term verification
 - Summary data initially
 - Can't independently verify or reproduce results
 - Not consistent with IQG

GLLA Data Management Challenges



- Quick turnaround data review does not give you the full data quality picture
 - Review as much of the quality information up front as possible
 - Standardized reporting facilitates this review
 - Moving toward Staged Electronic Data Deliverable (SEDD) for GLLA projects
- Independent review of all data and supporting information still needs to happen
 - Completeness checks
 - Review of data usability in the context of project goals

GLLA Data Management Objectives



1. Archive and maintain current and historical sediment confirmation data
2. House the associated quality assurance and quality control data for the projects
3. Maintain a simplified system for easy maintenance and ongoing support
4. Allow easy and user-friendly public access to the system
5. Minimize data manipulation required for upload given data are being received in a variety of formats (traceability)
6. Compatibility with NOAA's Query Manager system and the National Sediment Inventory (NSI)

Great Lakes Sediment Database



- Compatible with NOAA's Query Manager (QM) system and the National Sediment Inventory
- Meets Office of Management and Budget's Information Quality Guidelines for transparency
 - "Evidence of Proof"

GLLA Data Management



- Allowable Data Reporting Formats
 1. Query Manager Template (with QC)
 2. EPA Region V EDD
 3. Staged Electronic Data Deliverable (SEDD)
 - With field data in QM template or Region V EDD

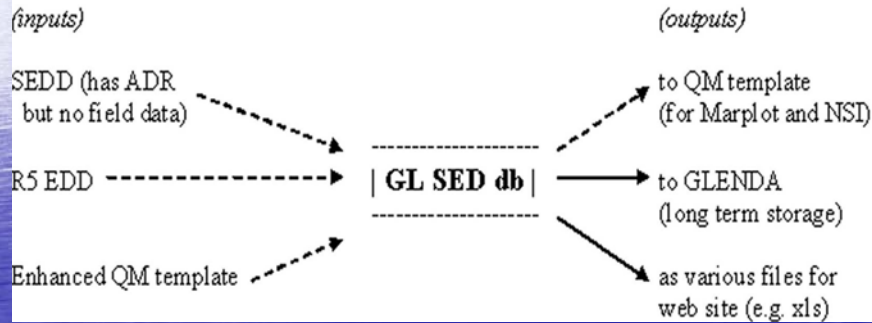
GLLA Data Management



- Data Reporting Guidelines
 - Stand alone document currently included as an attachment to the QMP
 - For submittal to contractors and laboratories responsible for reporting data under the Legacy Act
 - Contains data standard templates, references and user guides needed to report data



GLLA Data Management Strategy



SEDD

- Inter-agency effort to create a generic format for electronic delivery of analytical data for environmental programs
- The data deliverable generated by SEDD is an industry-standard Extensible Markup Language (XML) file
- A major advantage is that SEDD can be implemented in stages from summary data to complete reproducible data sets with all QC

Case Study Sediment Assessment



- Wisconsin sampling project
 - More than 400 samples
 - ERRS contractor
 - EPA wanted to use SEDD Stage 3, but also did not want to limit competition
 - Had to go with Stage 2a as not enough labs bid on SEDD Stage 3
 - Used ADR package
 - Still required manual review for about half of the required checks

Case Study Sediment Remediation



- Ashtabula River Sediment Remediation Project
 - RAC contractor conducted remediation
 - Contract Laboratory Program (CLP) labs used
 - Data turnaround very quick
 - Contract compliance screening (CCS) valuable and efficient
 - ESAT useful (good records management), assured raw data matched EXES review
 - Cost standardized
 - Merging with validation approach
 - Still awaiting lab QC data from RAC contractor, but have through Stage 3 CLP SEDD output

Case Study Sediment Assessment



- Riverview Michigan
 - More than 200 samples
 - reported in more than 60 files from multiple labs
 - One lab
 - EDD is an Excel spreadsheet designed to print nicely
 - Approximately 1 hour is needed per file to remove headers and transpose data to obtain workable data file
 - Validator assigned flags by marking up hard-copy summary reports
 - Contracted laboratories
 - Contractor responsible for data verification received SEDD 3 files, but provided SEDD 1 files, as they wanted to use ADR
 - QC data were not provided, "I did not know you wanted that level of QC data electronically"

Successful Data Management



- Black Lagoon project
 - Data quality objective table
 - Statistical basis for design
 - Data reporting standard
 - Data verification narratives
 - attached to database
 - includes majority of components of data usability report
 - Database
 - Final report

Recommendations



- Encourage adoption of SEDD
 - Few labs routinely reporting in SEDD outside of CLP
- Need to develop Non-proprietary automated data review software for 2a and 2b
- Standardize field data reporting for removal and emergency response projects
 - Incorporate into ESAT process
- Expand National Functional Guidelines qualifiers to provide information on QC failures
 - J is used to handle a wide variety of QC issues

Recommendations



- Expand tools to work with non-traditional methods
 - toxicity,
 - physical data,
 - toxicity characteristic leaching procedure (TCLP) for permits
- Provide for flexibility in analytical methodology
- Provide for project-specific measurement quality objectives
- Encourage data usability assessment in merged project data sets

Project Successes



- GLSED
 - Version 3 of GLSED will be completed in April, containing data for at least six projects
 - More than 65,000 sample results
 - Database also contains data verification narratives
- Working directly with Superfund Program great relationship with OSRTI-ASB, Regions 5 & 2
- SEDD and CLP significantly reduced costs and time to complete initial data review through EXES

“The Great Lakes Legacy Act Program has developed very comprehensive approaches for quality management that entail ‘cradle-to-grave’ QA requirements. These approaches adopt the best practices for sediment remediation and are endorsed by all stakeholders.”

from the EPA Headquarters Quality Management Review of GLNPO conducted in August 2006



EXES – An Advanced System for Environmental Data Assessment and Data Validation



Michael S. Johnson - USEPA Analytical Services Branch (ASB/OSRTI)

Nazy Abousaedi - Computer Sciences Corporation (CSC)

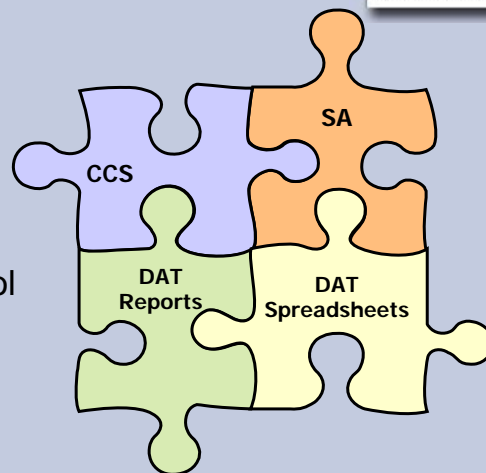
Yan Yang - Computer Sciences Corporation (CSC)

April 23, 2008

Electronic Data eXchange and Evaluation System (EXES) Processes and Deliverables



- Laboratory Self Assessment (SA)
- Contract Compliance Screening (CCS)
- Data Assessment Tool (DAT) Reports
- DAT Spreadsheet Reports



April 23, 2008

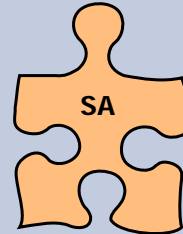
EXES- An Advanced Data Validation Tool/ Michael S. Johnson, Nazy Abousaedi & Yan Yang

2

Laboratory Self Assessment



- Allows laboratory to self inspect data prior to delivery to clients.
- Checks reporting and technical requirements.
- Provides detailed reports to the laboratory



April 23, 2008

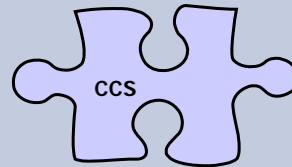
EXES- An Advanced Data Validation Tool/ Michael S. Johnson, Nazy Abousaeedi & Yan Yang

3

Contract Compliance Screening (CCS)



- Completeness and Compliance check of the electronic data.
- Based upon lab contract and method requirements.
- Flexible to accommodate variations to the analytical methods.
- Detailed report provided to the lab and summary report provided to the client.



April 23, 2008

EXES- An Advanced Data Validation Tool/ Michael S. Johnson, Nazy Abousaeedi & Yan Yang

4

Data Assessment Tool (DAT)



- Performs recalculation of information from raw data.
- Performs data validation checks based on national or data user's guidelines.
- Can be customized based upon client, project, or method variations.
- Designed to assist data validation, NOT replace.

April 23, 2008

EXES- An Advanced Data Validation Tool/ Michael
S. Johnson, Nazy Abousaeedi & Yan Yang

5

DAT Deliverables Reports



- Series of reports based upon validation criteria.
- Summary reports provide final results and validation flags.
- Results based upon data user's validation criteria.
- Received within 24-48 hours of data delivery from laboratory.



April 23, 2008

EXES- An Advanced Data Validation Tool/ Michael
S. Johnson, Nazy Abousaeedi & Yan Yang

6

DAT Deliverables Spreadsheets



- Customized for each user or project.
- Usually an Excel spreadsheet containing up to 60-70 fields.
- Amenable to loading into databases.
- Combines lab data, field data, and validation results.
- DAT Editor available to assist data validators.



April 23, 2008

EXES- An Advanced Data Validation Tool/ Michael
S. Johnson, Nazy Abousaeedi & Yan Yang

7

Hard Copy and Electronic Data Comparison



- Performed between recalculated results in spreadsheets and hard copy Form 1s.
- Is a spot check to catch possible systematic errors.
- If a problem is detected (unusual), client is notified and the lab is requested to rectify error.
- Performed on all DAT spreadsheets before distribution to users.

April 23, 2008

EXES- An Advanced Data Validation Tool/ Michael
S. Johnson, Nazy Abousaeedi & Yan Yang

8

For More Information



- Michael S. Johnson
 - Phone: 703-603-0266
 - Email: johnson.michaels@epa.gov
- Nazy Abousaeedi
 - Phone: 703-818-4233
 - Email: nazy@fedcsc.com
- Yan Yang
 - Phone: 703-818-4524
 - Email: yyang6@fedcsc.com

April 23, 2008

EXES- An Advanced Data Validation Tool/ Michael
S. Johnson, Nazy Abousaeedi & Yan Yang

9

Improving Efficiency in Regional Data Validation



Robert Runyon, USEPA Region 2

April 23, 2008

Overall Regional Approach



- Regional validation of all CLP data based on intended use.
- Use CLP electronic reports and spreadsheets to streamline validation process.
- Meet FASTAC turnaround objectives for data validation.
- Meet customer needs of timely data of known quality.

April 23, 2008

Data Validation Efficiency/ Robert Runyon

2

Use of Reports by the Regional Validators



- Use DAT validation reports to assist in identifying data issues, identifying affected samples, result qualification.
- Utilization of selected text in the DAT reports to generate final Regional data validation report.
- Allows validators to focus more on areas requiring professional judgment and site specific issues.

April 23, 2008

Data Validation Efficiency/ Robert Runyon

3

Use of the Spreadsheets



- Region 2 spreadsheets are designed for loading data into the Regional site database.
- The DAT editor facilitates Regional edits to the spreadsheet to incorporate any changes resulting from the Regional manual validation process.

April 23, 2008

Data Validation Efficiency/ Robert Runyon

4

Use of the Spreadsheets (Cont.)



- Printouts from the DAT editor are utilized for “human viewable” documentation.
- The final spreadsheet with the validated results is forwarded to the RPM and loaded into the Regional database.

April 23, 2008

Data Validation Efficiency/ Robert Runyon

5

Regional Efficiencies



- The DAT deliverables have reduced Regional validation times for organic data from approximately 2-3 hours/sample analysis in a full manual review to approximately .3-.75 hours/sample analysis for DAT-assisted review.
- Inorganic review times have been reduced from approximately .4-.7 hours/sample analysis to approximately .3-.4 hours/sample analysis.

April 23, 2008

Data Validation Efficiency/ Robert Runyon

6

Regional Efficiencies (Cont.)



- Estimated FY07 savings for DAT assisted CLP data validation was \$958,700 (organic \$870,400 and inorganic \$88,300) when compared to fully manual data validation.
- These efficiencies do not include the savings realized by the ability to load fully validated data into site databases.

April 23, 2008

Data Validation Efficiency/ Robert Runyon

7

For More Information



- Robert Runyon
 - Phone: 732-321-6645
 - Email: Runyon.Robert@epa.gov

April 23, 2008

Data Validation Efficiency/ Robert Runyon

8

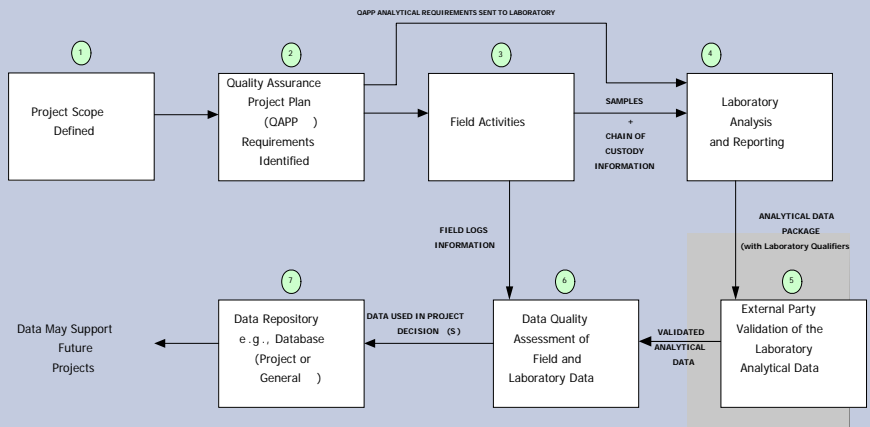
Improving Communication About Superfund Data Validation



Bruce Means
Chief, Analytical Services Branch

April 23, 2008

Typical Superfund Data Generation and Review Process



April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

2

Current State



- Often laborious, complex procedures.
- Inconsistency/ambiguity in:
 - terminology
 - guidance
 - practice
- Many different reviewing organizations.
- Communication about specifics of reviews inconsistent, incomplete.

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

3

Workgroup Goals



- Improve communication within Superfund about scope and content of lab analytical data verification and validation.
- Encourage appropriate use of data for
 - Task at hand
 - Future decisions

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

4

Approach Taken



- Facilitate communication through the use of “labels” that summarize verification and validation checks.
- Checks would be grouped into stages.
- Each stage to build on the checks from previous stage.
- Describe nature of review process (manual and/or electronic)

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

5

Range of Validation Checks



- Completeness
- Compliance
 - Sample-related QC
 - Instrument-related QC
- Recalculation
- Instrument output review

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

6

Completeness Checks



- To make sure that the requested data deliverables are provided.
- To determine that data requested are actually present in the deliverables.
- Can include hard copy and/or electronic formats.

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

7

Compliance Checks



- To compare analytical Quality Control (QC) results with the acceptance criteria, requirements or guidelines present in the project-specific Quality Assurance Project Plan (QAPP), regional data validation documents, analytical method(s) or contract.
 - *Sample-Related QC* (e.g., blank contamination, surrogate recoveries)
 - *Instrument-Related QC* (e.g., instrument calibration, tune)

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

8

Recalculation Checks



- The laboratory reported values (e.g., sample results, instrument calibration results) are verified by recalculation using instrument output data reported by the laboratory.
- Confirms that correct formulae and values were used in calculation of results.

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

9

Instrument Output Checks



- Actual instrument outputs should be checked to ensure that the laboratory reported analytes have been correctly identified and quantitated (e.g., are mass spectra properly identified? Are peak integrations correct?).

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

10

Proposed Validation Stage Labels



- Completeness Stage 1
- ...and Sample-related QC Stage 2a
- ...and Instrument-related QC Stage 2b
- ...and Recalculations Stage 3
- ...and Instrument output Stage 4

Possible Process Labels



- Electronic review only EO
- Manual validation M
- Electronic and manual EM

Possible Applications ?



- Sample Delivery Group (SDG) level.
- Analyte-specific codes possible.

Example: A verification and validation conducted using both electronic tools and manual expertise to examine completeness and compliance checks would be labeled:

Stage 2b Validation, Electronic and Manual
(example label code: S2bVEM)

Desired Outcome



- Third party reviewers associate the reviewed data with its review stage as data is shared with decision makers.
- Data users / decision makers quickly recognize the nature of review performed on data prior to use.
- Future use of data is facilitated by labels that travel with data.

Additional Information



Bruce Means

Chief, Analytical Services Branch
Technology and Field Services Division
Office of Superfund Remediation and
Technology Innovation

means.bruce@epa.gov

703-603-8815

April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

15

Questions?



April 23, 2008

Improving Communication About Superfund Data
Validation/ Bruce Means

16