Electronic Data Assessment

Sara Goehl
EPA Office of Superfund Remediation and Technology Innovation
Analytical Services Branch

Presentation Overview

- Electronic Data Assessment
- What Electronic Assessment Does and Doesn't Do
- The Contract Laboratory Programs Approach to Electronic Data Assessment
- Enhanced EXES Overview and Live Demo

Poll Question

What is your role in the data assessment process?

a) EPA Project Officer

b) Data Validator

c) EPA Superfund Field Contractor

d) Association with other Federal, State or Private Organization

e) Other

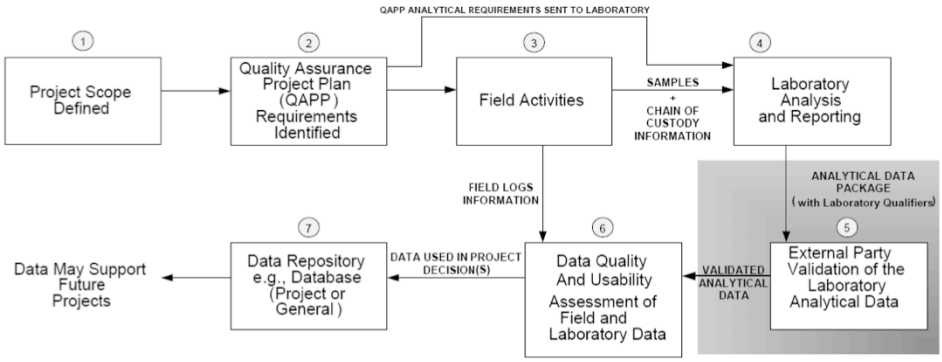
Poll Question

Do you perform validation on or otherwise work with EPA Contract Laboratory Program data?

a) Yes

b) No

FIGURE 1: TYPICAL SUPERFUND DATA GENERATION AND REVIEW PROCESS



From *Guidelines for Labeling External Data for Superfund Use (EPA 2009)*

Electronic Data Assessment

- Electronic Analytical data (.xml, .csv, etc.) is evaluated against pre-defined criteria using a software program
- The software program may recalculate data, apply flags, and generate a report back the user identifying any non-compliance or performance issues
- Data assessment is only as good as the data input (Garbage in = Garbage out)

Electronic Data Assessment - Advantages

- Saves time (and $$)
- Increases efficiency
- Speeds delivery to end user
- Consistent evaluation
- Can quickly and easily evaluate entire data set rather than a subset
- Can often generate spreadsheets that can be directly uploaded to databases

Electronic Data Assessment - Disadvantages

- Requires very specific formatting of files
- Software output from analytical instrument may not be compatible with formatting requirements
- Different organizations (EPA, DOD, States, etc.) require different outputs for database compatibility
- Most data review programs are not flexible, users cannot evaluate anything beyond preprogrammed checks

Staged Electronic Data Deliverables (SEDD)

- Uniform format for the electronic delivery of environmental analytical data developed for EPA and USACE
- Stages refers the level of data included in the SEDD file
- SEDD Specifications identifies required information and valid values

Valid Values – Value entered into a specific field that can be recognized by the software

SEDD Benefits

Currently labs produce over 100 different types of electronic deliverables to meet the requirements of different Federal Agencies, States, and other customers

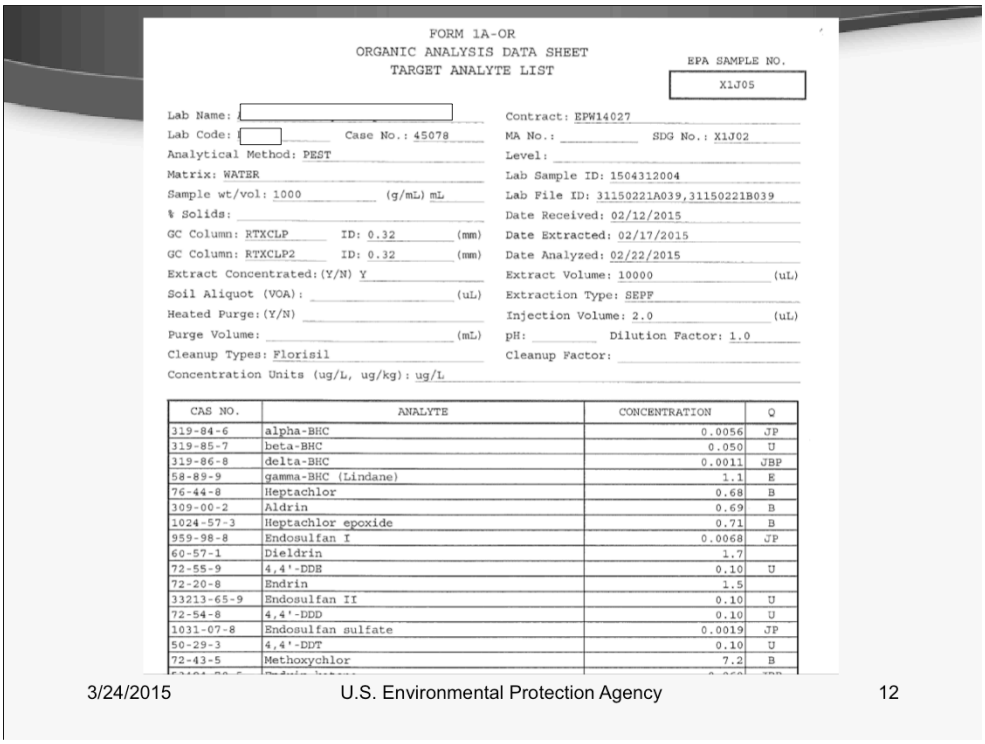
- The SEDD specifications establishes a uniform electronic format that can meet the needs of multiple agencies and programs
- Provides a common set of valid values for methods, analytes, units, etc.
- Standard format enables automated review

SEDD Stages

SEDD Stage 1

- Contains the minimum number of analytical data elements to convey results ONLY to the end data user.

- Equivalent to CLP Form 1



FORM 1A-OR
ORGANIC ANALYSIS DATA SHEET
TARGET ANALYTE LIST

EPA SAMPLE NO.
X1J05

Lab Name: [redacted]
Contract: EPW14027
Lab Code: [redacted] Case No.: 45078 MA No.: _____ SDG No.: X1J02
Analytical Method: PEST
Level: _____
Matrix: WATER
Lab Sample ID: 1504312004
Sample wt/vol: 1000 (g/mL) mL
Lab File ID: 31150221A039,31150221B039
% Solids: _____
Date Received: 02/12/2015
GC Column: RTXCLP ID: 0.32 (mm)
Date Extracted: 02/17/2015
GC Column: RTXCLP2 ID: 0.32 (mm)
Date Analyzed: 02/22/2015
Extract Concentrated: (Y/N) Y
Extract Volume: 10000 (uL)
Soil Aliquot (VOA): _____ (uL)
Extraction Type: SEPF
Heated Purge: (Y/N) _____
Injection Volume: 2.0 (uL)
Purge Volume: _____ (mL)
pH: _____ Dilution Factor: 1.0
Cleanup Types: Florisil
Cleanup Factor: _____
Concentration Units (ug/L, ug/kg): ug/L

CAS NO.	ANALYTE	CONCENTRATION	Q
319-84-6	alpha-BHC	0.0056	JP
319-85-7	beta-BHC	0.050	U
319-86-8	delta-BHC	0.0011	JBP
58-89-9	gamma-BHC (Lindane)	1.1	E
76-44-8	Heptachlor	0.68	B
309-00-2	Aldrin	0.69	B
1024-57-3	Heptachlor epoxide	0.71	B
959-98-8	Endosulfan I	0.0068	JP
60-57-1	Dieldrin	1.7	
72-55-9	4,4'-DDE	0.10	U
72-20-8	Endrin	1.5	
33213-65-9	Endosulfan II	0.10	U
72-54-8	4,4'-DDD	0.10	U
1031-07-8	Endosulfan sulfate	0.0019	JP
50-29-3	4,4'-DDT	0.10	U
72-43-5	Methoxychlor	7.2	B

SEDD Stages

SEDD Stage 2a

- Builds on SEDD Stage 1 and includes Method/ Sample-Specific QC Information
  - Laboratory Control Samples
  - Blanks
  - Spike Samples
  - Duplicates
  - Surrogate Recoveries

SEDD Stages

SEDD Stage 2b

- Builds on SEDD Stage 2a and includes Instrument QC data
  - Mass Spec Tune Data
  - Relative Response Factors
  - Calibration Information
  - Peak Information

SEDD Stages

SEDD Stage 3

Builds on SEDD 2b and adds additional instrument measurement data to allow for the recalculation of reported results.

Contract Laboratory Program

SOM01.X and SOM2.x -SEDD 3

ISM01.X –SEDD 2b

ISM02.X –SEDD 3

SEDD Stages

SEDD Stage 4

Builds on SEDD 3 and adds instrument raw data files.

- Instrument raw data files are included in CLP hardcopy or pdf data package, but not in the electronic SEDD file
- No plans to move to SEDD Stage 4 for CLP

SEDD Resources

General Information:
http://www.epa.gov/fem/sedd-geninfo.htm

SEDD 5.2 Specifications:
http://www.epa.gov/fem/seddspec52.htm

LIMS Data Capture and LIMS/SEDD Relationships:
http://www.epa.gov/fem/pdfs/lims-data-capture.pdf
http://www.epa.gov/fem/pdfs/lims-sedd.pdf

Combining Electronic Assessment with Data Validation

Data Verification – *Process for evaluating data completeness to confirm that all data requested from the laboratory has been received and complies with the specified requirements*

Data Validation- *Process for evaluating analyte and sample specific compliance of laboratory data with analytical methods, procedures, and contractual requirements*

Data Usability- *Evaluation of the results of the data verification, validation, field data assessment, and other factors to determine if data meets Data Quality Objectives and is suitable for the intended decision*

* *Definitions given are in context of and for purposes of this presentation*

Data Validation

Analytical Data Verification and Validation Tiers

Tier 1 Verification and Validation (Stage 1, Level 1)

- Assessment is based only on the completeness and compliance of sample receipt conditions.

Tier 2a Verification and Validation

- Assessment is based on completeness and compliance of sample receipt conditions (Tier 1) AND sample-related QC results.

From *Guidelines for Labeling External Data for Superfund Use (EPA 2009)*

Data Validation

Analytical Data Verification and Validation Tiers

Tier 2b Verification and Validation

- Assessment is based on completeness and compliance of sample receipt conditions (Tier 1) AND sample-related QC results (Tier 2a) AND instrument related QC results should be considered a Tier 2b level validation.

Tier 3 Verification and Validation

- Assessment is based on completeness and compliance of sample receipt conditions (Tier 1) , sample-related QC results (Tier 2a), instrument related QC results (Tier 2b) AND recalculation checks should be considered a Tier 3 level validation.

From *Guidelines for Labeling External Data for Superfund Use (EPA 2009)*

Data Validation

Analytical Data Verification and Validation Tiers

Tier 4 Validation and Verification

- Assessment is based on completeness and compliance of sample receipt conditions (Tier 1) , sample-related QC results (Tier 2a), instrument related QC results (Tier 2b), recalculation checks (Tier 3) AND a review of the actual instrument output should be considered a Tier 4 validation.

Electronic Assessment is currently only available for up to Tier 3 (based on SEDD level).

Data Validation

SEDD Stage ≠ Data Validation Level

- Data validation approach should be based on the Data Quality Objectives developed at the beginning of the project
- SEDD Stage will be based on method and laboratory capability
- Electronic assessment will be based on available software

Setting an Approach to Data Validation

Electronic Data Assessment is based on the software available for the method and the SEDD level that the lab is able to produce.

- CLP –SEDD Stage 2b or 3
- Non-CLP labs may be able provide a 2a
- Limited capability for electronic assessment of .csv or EXCEL files

Data Validators should develop data review SOPs in accordance with project Data Quality Objectives, Regional requirements, and SEDD checker specifications.

Data Validation

Analytical Data Verification and Validation Tiers

- Using higher stages of analytical verification and validation does not typically result in higher data quality
- Rather, the quality of the analytical data becomes more transparent as more stages of verification and validation are conducted
- As a result, the usability of the analytical data for its intended use becomes more apparent

From *Guidelines for Labeling External Data for Superfund Use (EPA 2009)*

Data Validation

Guidelines for Labeling Externally Validated Data for EPA Superfund Use (EPA 2009)

- Gives a more thorough overview of the level of review expected for each tier
- Lists the minimum checks recommend for each data validation tier
- Provides recommended terminology for communicating to data recipients and other data users the steps and manner used for analytical data verification and validation

From *Guidelines for Labeling External Data for Superfund Use (EPA 2009)*

Data Validation

Recommended Terminology and Labels for Communicating the Stages and Processes Used for Laboratory Analytical Data Verification and Validation

Label Corresponding Label Code	Label Corresponding Label Code
Stage_1_Validation_Electronic	S1VE
Stage_1_Validation_Manual	S1VM
Stage_1_Validation_Electronic_and_Manual	S1VEM
Stage_2A_Validation_Electronic	S2AVE
Stage_2A_Validation_Manual	S2AVM
Stage_2A_Validation_Electronic_and_Manual	S2AVEM
Stage_2B_Validation_Electronic	S2BVE
Stage_2B_Validation_Manual	S2BVM
Stage_2B_Validation_Electronic_and_ Manual	S2BVEM
Stage_3_Validation_Electronic	S3VE
Stage_3_Validation_Manual	S3VM
Stage_3_Validation_Electronic_and_Manual	S3VEM
Stage_4_Validation_Electronic	S4VE
Stage_4_Validation_Manual	S4VM
Stage_4_Validation_Electronic_and_ Manual	S4VEM
Not_Validated	NV

From *Guidelines for Labeling External Data for Superfund Use (EPA 2009)*

27

The Contract Laboratory Program's Approach to Electronic Data Validation

Contract Lab Program

- National network of commercial laboratories that provide analytical services to EPA Superfund clients
- Serves all ten EPA Regions and Brownfields
- All labs must follow the same analytical requirements (Statement of Work). CLP does not allow “deviations”
- CLP uses electronic and manual screening to evaluate contract compliance and data validation
- Sanctions are applied for late and noncompliant data

EXES

The CLP streamlined the requirements for SEDD deliverables, and developed **EXES** (the Electronic Data eXchange and Evaluation System), for fast, efficient, and standardized data assessment.

- EXES is a software tool designed to quickly and thoroughly evaluate laboratory analytical data
- EXES performs a series of “tests” on electronic data to identify potential problems and flags data quality or usability issues
- Provides users with reports identifying potential data quality issues
- Marries data from Scribe or other field data input and produces a standard or customized electronic deliverable

U.S. Environmental Protection Agency

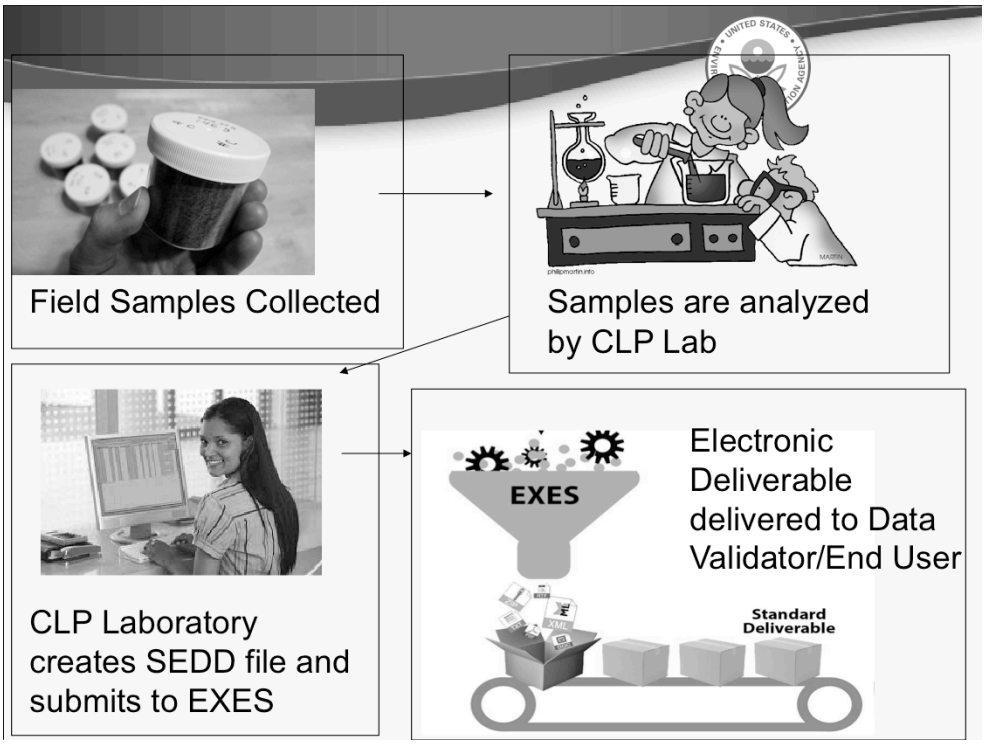
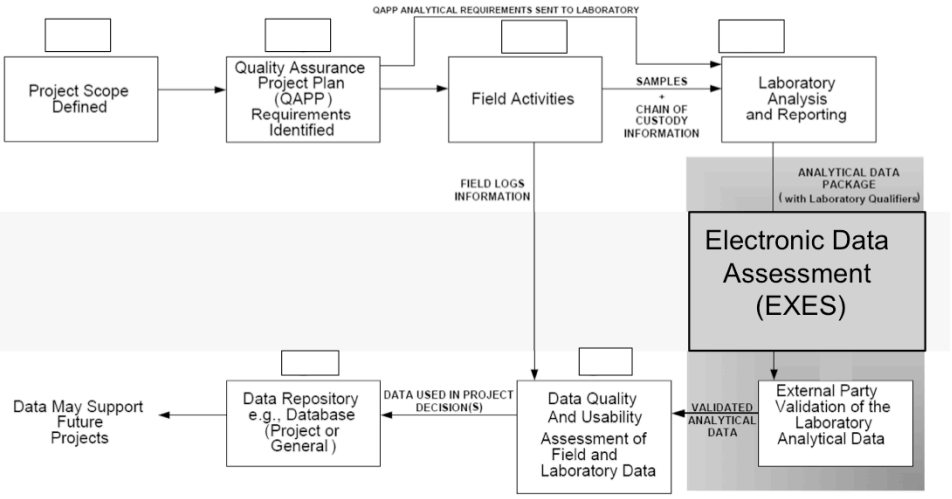

Field Samples Collected
Samples are analyzed by CLP Lab
CLP Laboratory creates SEDD file and submits to EXES
EXES
Electronic Deliverable delivered to Data Validator/End User
Standard Deliverable

FIGURE 1: TYPICAL SUPERFUND DATA GENERATION AND REVIEW PROCESS



EXES

The EXES Electronic Data Assessment Consists of three parts:

Self-Inspection- The laboratory is allowed to “check” their SEDD files for compliance and correct any missing/incorrect data before final submission to EPA.

Initial Assessment- The EXES software evaluates the file to make sure it contains a minimum amount of information and that information is in the correct format.

Full Assessment- EXES performs the full suite of electronic checks and provides a report back to the laboratory.

EXES

Once the SEDD file has been submitted to EPA

Contract Compliance Screening (CCS) – This report evaluates whether the laboratory correctly adhered to the Statement of Work (analytical method). Laboratories may receive sanctions for not following contract requirements.

National Functional Guidelines (NFG) Evaluation- this evaluation compares analytical results to data validation guidelines outlined in the “National Functional Guidelines for Superfund Organic/Inorganic Data Review”. Laboratories are not penalized for NFG defects.

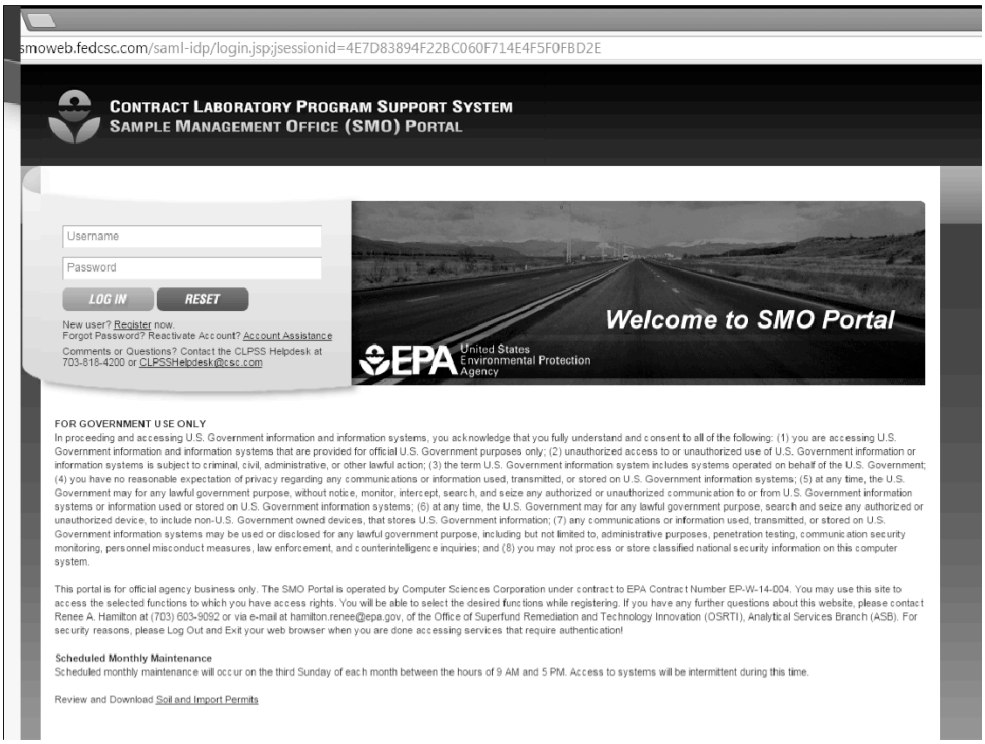
What About the Data Package?

- Labs are required to submit a hardcopy of all required CLP Forms, case narratives, chain of custody information, and instrument data to the EPA Region as well as a pdf copy to the EPA Region and SMO
- The Sample Management Office performs a screen of the pdf for completeness
- SMO does not compare values in hardcopy against values in EDD

CLP Data Verification/Validation

The hardcopy received by the EPA Regions is considered the “Official Copy” of the data. Regions should ensure this data package contains all necessary information (data verification)

EPA Regions can access a pdf of the data package, Universal Data Deliverable, Regional Specific Spreadsheets, and EXES CCS and NFG Reports via “Electronic Data Manager” (EDM) in the SMO Portal


smoweb.fedcsc.com/saml-idp/login.jsp;jsessionid=4E7D83894F22BC060F714E4F5F0FBD2E
CONTRACT LABORATORY PROGRAM SUPPORT SYSTEM
SAMPLE MANAGEMENT OFFICE (SMO) PORTAL
Username
Password
LOG IN
RESET
New user? Register now.
Forgot Password? Reactivate Account? Account Assistance
Comments or Questions? Contact the CLPSS Helpdesk at 703-818-4200 or CLPSSHelpdesk@csc.com
Welcome to SMO Portal
EPA United States Environmental Protection Agency
FOR GOVERNMENT USE ONLY
In proceeding and accessing U.S. Government information and information systems, you acknowledge that you fully understand and consent to all of the following: (1) you are accessing U.S. Government information and information systems that are provided for official U.S. Government purposes only; (2) unauthorized access to or unauthorized use of U.S. Government information or information systems is subject to criminal, civil, administrative, or other lawful action; (3) the term U.S. Government information system includes systems operated on behalf of the U.S. Government; (4) you have no reasonable expectation of privacy regarding any communications or information used, transmitted, or stored on U.S. Government information systems; (5) at any time, the U.S. Government may for any lawful government purpose, without notice, monitor, intercept, search, and seize any authorized or unauthorized communication to or from U.S. Government information systems or information used or stored on U.S. Government information systems; (6) at any time, the U.S. Government may for any lawful government purpose, search and seize any authorized or unauthorized device, to include non-U.S. Government owned devices, that stores U.S. Government information; (7) any communications or information used, transmitted, or stored on U.S. Government information systems may be used or disclosed for any lawful government purpose, including but not limited to, administrative purposes, penetration testing, communication security monitoring, personnel misconduct measures, law enforcement, and counterintelligence inquiries; and (8) you may not process or store classified national security information on this computer system.
This portal is for official agency business only. The SMO Portal is operated by Computer Sciences Corporation under contract to EPA Contract Number EP-W-14-004. You may use this site to access the selected functions to which you have access rights. You will be able to select the desired functions while registering. If you have any further questions about this website, please contact Renee A. Hamilton at (703) 603-9092 or via e-mail at hamilton.renee@epa.gov, of the Office of Superfund Remediation and Technology Innovation (OSRTI), Analytical Services Branch (ASB). For security reasons, please Log Out and Exit your web browser when you are done accessing services that require authentication!
Scheduled Monthly Maintenance
Scheduled monthly maintenance will occur on the third Sunday of each month between the hours of 9 AM and 5 PM. Access to systems will be intermittent during this time.
Review and Download Soil and Import Permits

What is EXES Data Manager (EDM)?

EDM provides a Web-based/centralized data management tool:

- Stores all laboratory deliverables and data assessment and NFG evaluation reports
- Provides spreadsheets and outputs including both field and analytical data ready for upload to databases and mapping tools
- Allows users to edit and regenerate all dynamic reports
- Provides an audit tracking log of all changes made by user name and date/time
- Who can have access to what portion of data or reports in EDM can be set based on data user's request

Universal Data Deliverable

- Data is provided to the EPA Regions in the **Universal Data Deliverable**
- Goal of Universal Data Deliverable is to improve consistency and reduce project-specific formats
- Most Regions also get Region specific formats
  - Equis-compatible, tab-delimited, etc.
- EPA Regions do not receive SEDD files through Electronic Data Manager (EDM)

CLP Data Verification/Validation

Once the data package and electronic deliverable arrive at the Region the SEDD and pdf have underwent the following data validation levels:

- SOM01.X SEDD: Electronic Tier 3 based on National Functional Guidelines (S3VE)
- ISM01.X: Electronic Tier 2b based on National Functional Guidelines (S2BVE)
- SOM02.X: Electronic Tier 3 based on National Functional Guidelines (S3VE)
- Data Package for all SOWs (both ISM and SOM): Tier 1-completeness only (S1VM)

Regional Role in CLP Data Verification/ Validation

The EPA Regions approach to data validation of CLP data varies by Region and Project-specific Data Quality Objectives (DQOs)

- Data Validation approach should outlined in the QAPP
- Should include data verification of the hardcopy at a minimum (Tier 1)
- Recommend ensuring values reported on CLP Form 1s match values reported in the EDD
- Many Regions use EXES generated NFG Reports to develop a “Tiered-Approach” to data validation

Regional Role in CLP Data Verification/ Validation

Many Regions use NFG Reports provided by EXES to develop a “Tiered-Approach” to data validation

- All data packages receive a “completeness check”/ data verification (Tier 1 S1VM).
- Data packages where EXES has identified the most “issues” receive the most scrutiny (Tier 3 or Tier 4)
- Random quality checks on all other data packages (Tier 2 or Tier 3)

This general approach should be outlined in the QAPP and SOPs used by validator

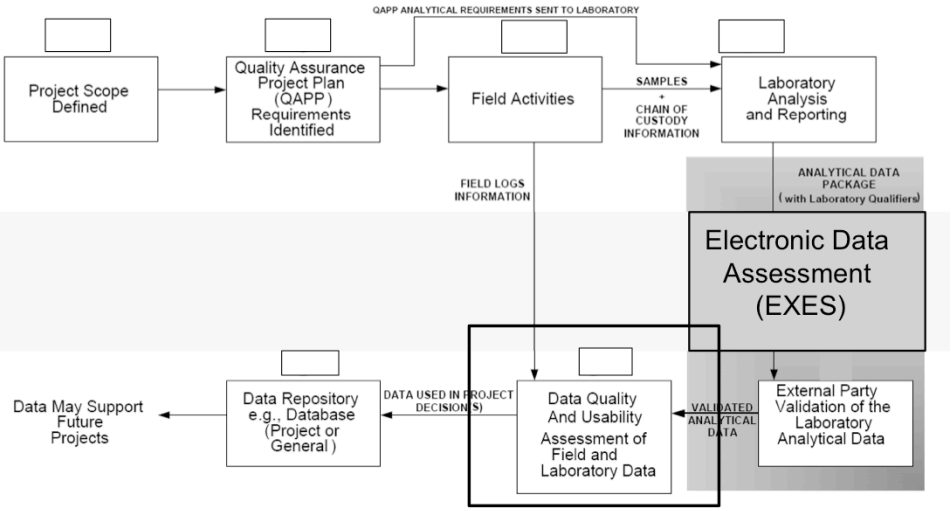
Advantages of a "Tiered Approach"

- Reduces time it takes for data validation and speeds delivery to end user
- Allows EPA Project Officer and Data Validators to use professional judgment rather than a "one size fits all" approach
- Approach can be modified based on project-specific needs
- EPA ASB can help EPA Regions develop a tiered approach for specific projects or Region-wide based on EXES evaluation

Regional Role in CLP Data Verification/ Validation

- EPA Regions are responsible for working with their CLP COR if more information is needed from CLP the laboratory to complete data validation
- EPA Regions should report any issues of repeated non-compliance or potential fraud to their EPA CLP COR and the EPA Analytical Services Branch
- EPA Regional Data Validators and Project Officers should contact the SMO helpdesk with any issues accessing the SMO Portal

FIGURE 1: TYPICAL SUPERFUND DATA GENERATION AND REVIEW PROCESS



A Note on Field Data

- SCRIBE is the preferred field data collection program for EPA Superfund
- CLP Samplers are required to submit field data information to SMO via an xml file within 72 hours of sample shipment
- CLP Laboratories do not see the field information. EXES combines field and lab information into the final electronic deliverable
- EXES currently does not perform checks on the field data

Field Data Validation

The QAPP should outline the Tier of analytical data assessment needed to meet Data Quality Objectives. *However,* data should not be considered usable for a project until the field data has been evaluated.

- Does the data correspond to historical data at the site
- Is there evidence of field contamination (field blanks)
- How homogenous were the samples (field duplicates)
- Do the GPS coordinates match and make sense?

Data Usability

The Project Team and Regional QA are ultimately responsible for evaluating data usability

- Does the data meet the project Data Quality Objectives outlined in the QAPP?
- Do the Data Validation and Verification findings show that my data is of sufficient quality for the intended decision?
- How does my data compare to historical information?
- How do non-detects and method detection limits affect my decision?
- Do I need any more information to be able to make my decision?

Enhanced EXES

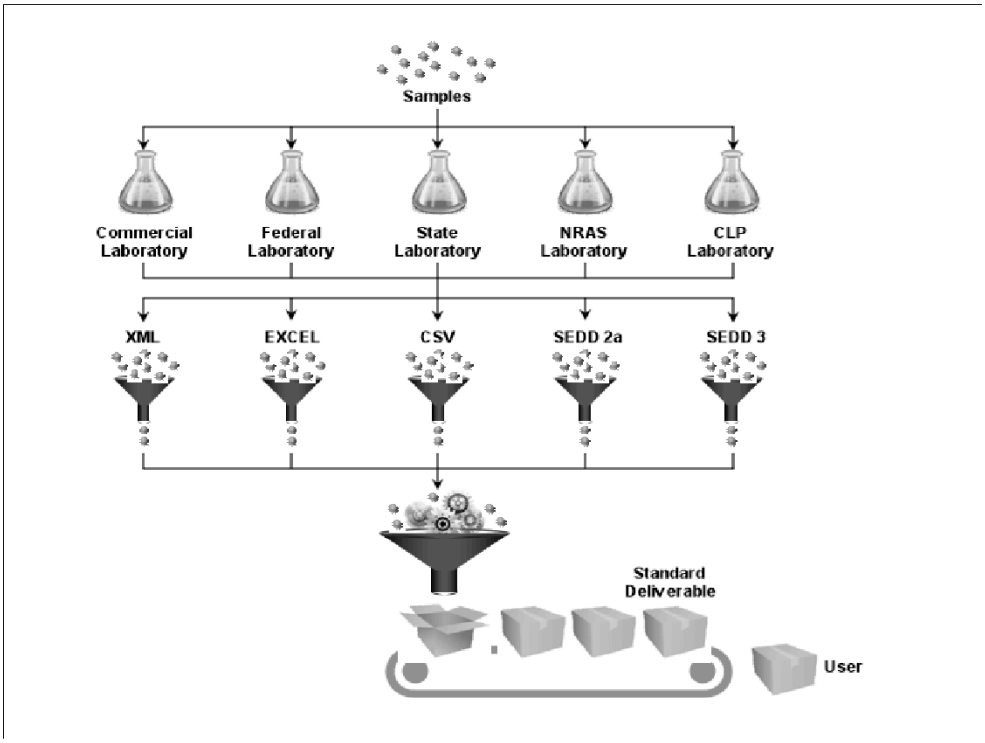
Enhanced EXES

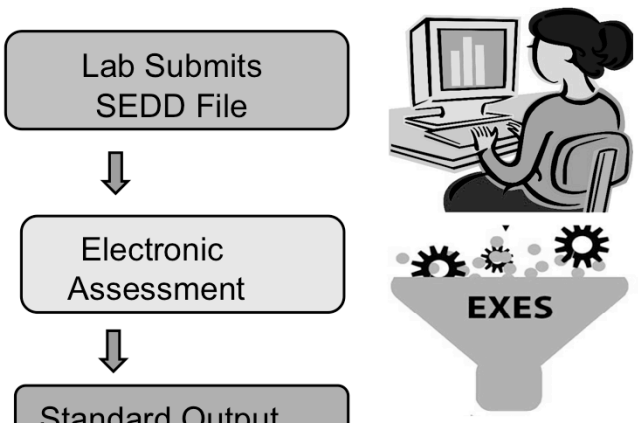
The EPA CLP Program will begin using new Statements of Work in Summer 2015 (ISM02.X and SOM02.x) and 2016 (HRSM01.X)

- SEDD files submitted under the new SOW's will be assessed using "Enhanced" EXES
- The EPA Analytical Services Branch is beginning pilot programs to add more methods to EXES and expand the use of EXES outside the CLP

What Makes Enhanced EXES Different?

- EXES users can now **create** new tests, **modify** existing tests, and create tests for **entirely new** methods without any computer programming.
- Methods are **reusable** and **customizable**, and can be shared among different users.
- EXES puts chemists in control of electronic data validation, computer programming is not required.
- EXES is **scalable** and **flexible**, and can be modified to meet the needs of any analytical program large or small.

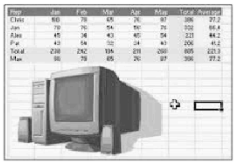

Samples
Commercial Laboratory
Federal Laboratory
State Laboratory
NRAS Laboratory
CLP Laboratory
XML
EXCEL
CSV
SEDD 2a
SEDD 3
Standard Deliverable
User



- Lab provides one electronic format
- EXES performs Standard or customized assessment
- User gets format they want

Excel Spreadsheet

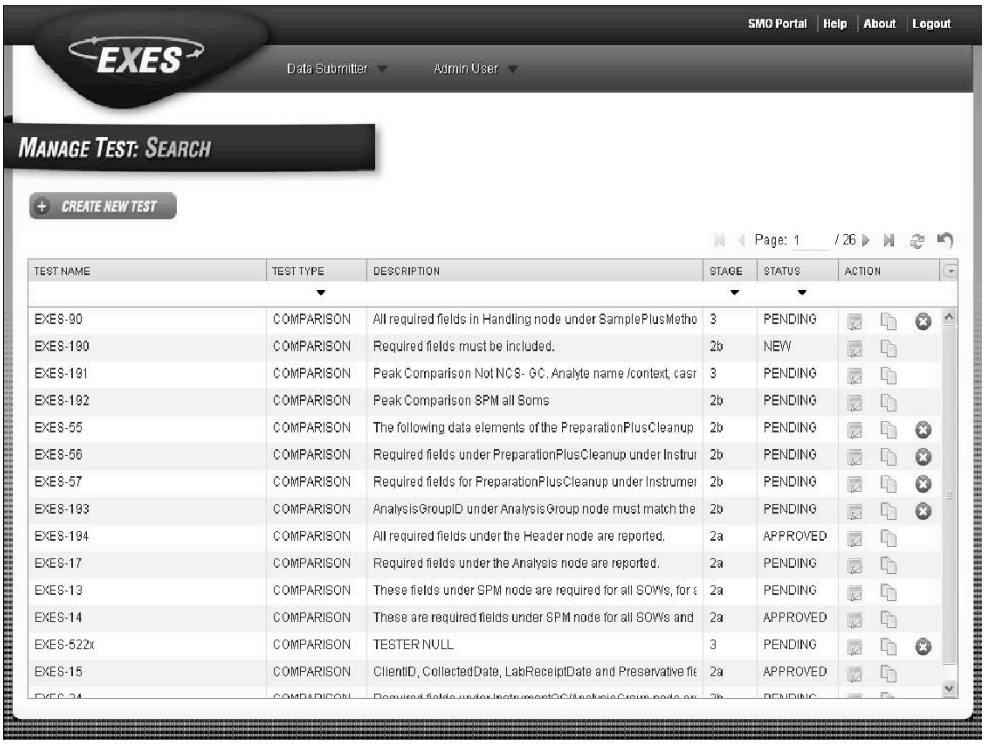
EQuIS-compatible

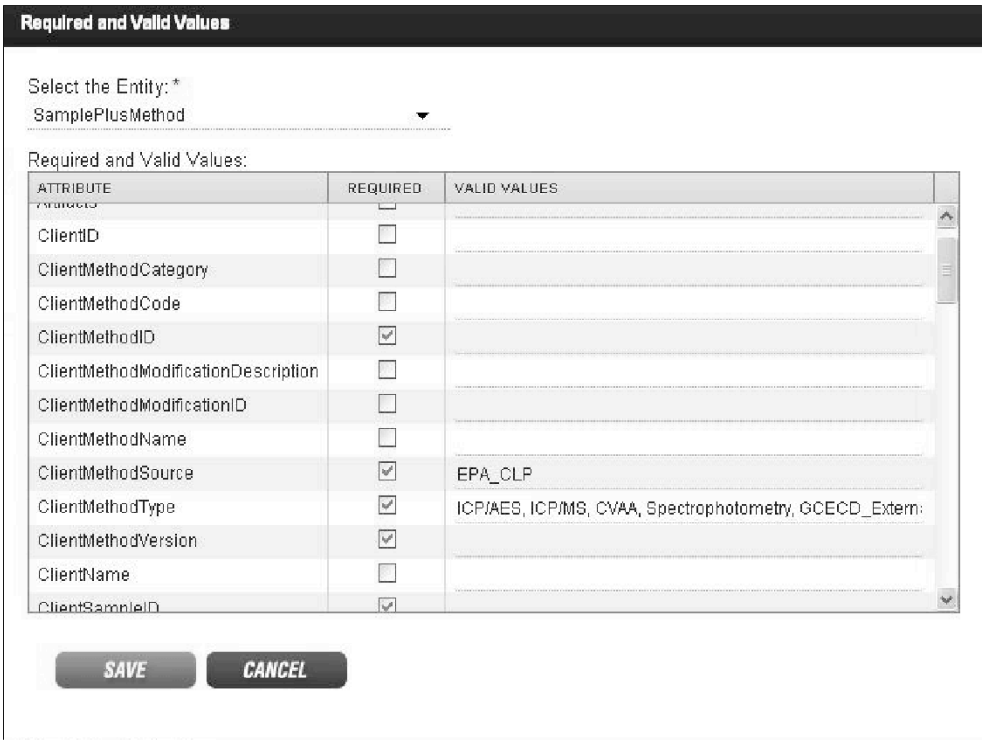


Other Electronic Format


SMO Portal | Help | About | Logout
EXES
Data Submitter
EXES DATA SUBMITTER
The EXES Data Submitter is your one stop location for submitting data. Use the entries below to submit your data and view queues.
Submit Data
View Queue
EXES
Contact


Add Filter
LEFT VALUE
Attribute
Condition Attribute
Function
Function: * roundValue
PARAMETERS
NAME
VALUE
originalValue
Peak.RRF
precisionType
Decimal
precisionNumber
3
Function Attribute
Group
OPERATOR
Equals
Not Equals
Greater Than
Less Than
Greater Than or Equal To
Less Than or Equal To
In List
Not In List
Between
Not Between
Is NULL
Is Not NULL
This is a Final Filter
RIGHT VALUE
Single Value
Attribute
Condition Attribute
Function
Function: * roundCalValue
Multiplier:
PARAMETERS
NAME
VALUE
in_edd_id
Header.NodeId
in_node_id
Peak
in_calc_name
RelativeResponseFactor
precisionType
Decimal
precisionNumber
3
Function Attribute
SAVE
CANCEL


SMO Portal
Help
About
Logout
EXES
Data Submitter
Admin User
MANAGE TEST: SEARCH
CREATE NEW TEST
Page: 1 / 26
TEST NAME
TEST TYPE
DESCRIPTION
STAGE
STATUS
ACTION
EXES-90 COMPARISON All required fields in Handling node under SamplePlusMetho 3 PENDING
EXES-100 COMPARISON Required fields must be included. 2b NEW
EXES-101 COMPARISON Peak Comparison Not NCS- GC. Analyte name /context, casr 3 PENDING
EXES-102 COMPARISON Peak Comparison SPM all Soms 2b PENDING
EXES-55 COMPARISON The following data elements of the PreparationPlusCleanup 2b PENDING
EXES-56 COMPARISON Required fields under PreparationPlusCleanup under Instrur 2b PENDING
EXES-57 COMPARISON Required fields for PreparationPlusCleanup under Instrumer 2b PENDING
EXES-103 COMPARISON AnalysisGroupID under AnalysisGroup node must match the 2b PENDING
EXES-104 COMPARISON All required fields under the Header node are reported. 2a APPROVED
EXES-17 COMPARISON Required fields under the Analysis node are reported. 2a PENDING
EXES-13 COMPARISON These fields under SPM node are required for all SOWs, for a 2a PENDING
EXES-14 COMPARISON These are required fields under SPM node for all SOWs and 2a APPROVED
EXES-522x COMPARISON TESTER NULL 3 PENDING
EXES-15 COMPARISON ClientID, CollectedDate, LabReceiptDate and Preservative fie 2a APPROVED


Required and Valid Values
Select the Entity:*
SamplePlusMethod
Required and Valid Values:
ATTRIBUTE
REQUIRED
VALID VALUES
ClientID
ClientMethodCategory
ClientMethodCode
ClientMethodID
ClientMethodModificationDescription
ClientMethodModificationID
ClientMethodName
ClientMethodSource
EPA_CLP
ClientMethodType
ICP/AES, ICP/MS, CVAA, Spectrophotometry, GCECD_Extern
ClientMethodVersion
ClientName
SAVE
CANCEL

Questions?

Sara Goehl
Goehl.sara@epa.gov
703-603-9097