

# Setting the Standard: Developing and Communicating Best Practices for Environmental Data Management

Brooke Roecker, Kristen Ward,  
Chris Mickle, Sarah Wright  
& Shauna McKellar

# Agenda

---

- Overview of ICEDM, BMP and path forward
- Synergies with other organizations
- White paper overview
  - Data Management Plan
  - Valid Values Management
  - Laboratory EDDs
  - Historical Data Migration/Management
- Q&A / Open discussion

# Overview of ICEDM

---

- ICEDM – International Conference for Environmental Data Management
- Forum to discuss data management issues, trends and innovations
- Platform Neutral
- Content provided by environmental data management professionals

# Overview of ICEDM

---

- Founded in 2010
- ICEDM Leadership Team
  - Sarah Wright (ERM)
  - Dan Higgins (Haley & Aldrich)
  - Lacy Smith (Wilcox)
  - Chris Mickle (Cardno)

<http://www.icedm.net/>

# ICEDM Mission Statement

---

Provide a venue for environmental data managers to collaborate and develop standard approaches for data management while cultivating leadership and technical excellence.

# Synergies With Other Organizations

---

## ● ICEDM Core Values

- Expanding the community
- Leverage available information
- Communicate
- Educate
- Participate

## ● Organizations

- ACWI (Advisory Committee on Water Information)
- EDDM (Environmental Disasters Data Management)
- NELAC (National Environmental Laboratory Accreditation Conference)

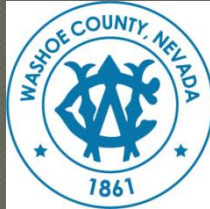
# ICEDM Best Management Practices Group

---

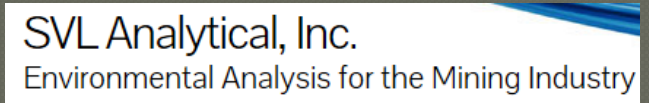
- Best Management Practices (BMP) Group formed in 2016
- Consists of volunteers from the Environmental Data Management (EDM) community
- Group mission: Provide vendor / platform neutral guidance



# 2016/2017 BMP Group Participants



Kennedy/Jenks Consultants  
Engineers & Scientists





# 2016 Best Practices Round Table Discussion

---

- Field Teams: Data Collection
- Legacy/Historical Data
- Standardizing Field Forms
- Security
- Data Deliverable Formats
- Define Terms
- Training
- Communication
- Team Definition
- Collection Method
- Audits
- Change Management/Audit Trail
- Needs Assessment
- Data Management Team
- Define Terminologies
- Data Management Plan Elements
- Data Management Basics
- Data Visualization

- Analytical:
  - Best Result
  - Qualifiers
  - Aggregated/Derived Data
  - Detected Data
- Valid Values
- Ethics
- Workflow
  - Defining Data System's Purpose
  - Validator
  - Laboratory
  - Field Crews
- Public Data Availability
- Data Verification Levels
- Nomenclature Guidance
- Maintaining query parameters for reporting
- Qualifiers/Reason Code

# 2016 Best Practices Round Table Discussion

- Field Teams: Data Collection
- **Legacy/Historical Data**
- Standardizing Field Forms
- Security
- **Data Deliverable Formats**
- Define Terms
- Training
- Communication
- Team Definition
- Collection Method
- Audits
- Change Management/Audit Trail
- Needs Assessment
- Data Management Team
- Define Terminologies
- **Data Management Plan Elements**
- Data Management Basics
- Data Visualization
- Analytical:
  - Best Result
  - Qualifiers
  - Aggregated/Derived Data
  - Detected Data
- **Valid Values**
- Ethics
- Workflow
  - Defining Data System's Purpose
  - Validator
  - Laboratory
  - Field Crews
- Public Data Availability
- Data Verification Levels
- Nomenclature Guidance
- Maintaining query parameters for reporting
- Qualifiers/Reason Code

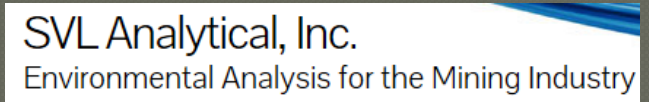
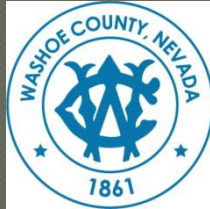
# ICEDM Best Management Practices Group

---

- BMP Lead (Kristen Ward – Langan)
- White paper categories chosen
  - Data Management Plan (Theresa Kennedy – ERM)
  - Valid Value Management (Chris Mickle - Cardno)
  - Historical Data Migration (Brooke Roecker - ddms)
  - Analytical/Laboratory EDDs (Karl Daines – SGS)
- BMP Group goal
  - Develop 4 white papers by May 2017 ICEDM
  - Open for public comment until 9/30/17

<http://www.icedm.net/icedm-bmp-group/>

# 2016/2017 BMP Group Participants



# Data Management Plan

---

## The Environmental Data Management Plan

Why do we need one and how do we get  
there?



# The Why and How

## Why do we need an EDMP?

### The Environmental Data Management Plan (EDMP)

#### Why do we need one and how do we get there?

*Everyone who has worked on an environmental project involving data collection has likely encountered the following scenario: needing a data set and, after finally discovering who has it and where, finding out it exists only in a spreadsheet, on a CD, in someone's desk drawer. The need for using or reporting the data then raises many questions, just a few of which may be: Are these all the data? Have the data been changed? What is the quality of the data? Who has ownership of the data? This paper will tell you why having an Environmental Data Management Plan (EDMP) is so important in addressing and avoiding the types of issues in this scenario, and will give you systematic guidance to writing your own EDMP. The actual process of writing your own EDMP shines a light on potential workflow issues, relative importance of different datasets, and communication gaps between data users, data generators, and data managers. The EDMP guides ongoing and future work, but its development with consideration to all aspects of environmental data management throughout a project's lifecycle is a benefit in and of itself. Also, the EDMP highlights that data have intrinsic value, beyond their initial purpose. This paper is presented in two parts: a paper describing why you need an EDMP, and an outline guiding the creation of an EDMP—how do you get there.*

## EDMP Template

### Environmental Data Management Plan (EDMP): Template and Content Guidelines

Date Created	[date]	[authors]
Date Revised	[date]	[notes]
Revision History	[date]	[notes]

#### 1 Introduction

##### 1.1 Purpose and Scope of the EDMP

This section should introduce the EDMP and note any key limitations in terms of scope to lead the reader to the appropriate section. For example, if the project is highly confidential with limitations to data use, note the limitation in this section and direct the reader to more information. Note also the scope in terms of period covered by the EDMP, and reference related documentation if not discussed in the sections below.

##### 1.2 Data Management Standards and Principles

The Data Management System (DMS) relies upon a consistent and logical framework to be laid out in the Environmental Data Management Plan (EDMP). Using a consistent framework allows growth and flexibility of the DMS. This framework provides a broadly similar, defined, and repeatable process that supports all data types. Defined logical workflows ensure that a process is traceable and repeatable from data acquisition to data reporting. Maintaining consistent and logical workflows allows the DMS to provide project and organization managers with efficient access to accurate information. Defining overarching standards and principles helps guide the remainder of process of creating the EDMP.



# Why Need an EDMP?

---

## ◉ Defines

- Standard structure, content, and format of data
- Practices and policies
- Roles and responsibilities

## ◉ Ensures

- Data Quality Objectives (DQOs) are met
- Relevant analysis and decision making
- Scientific, regulatory, and legal defensibility
- Preservation of inherent intrinsic value

## ◉ Improves

- Performance, accuracy, and efficiency

## ◉ Supports

- Effective dissemination of data

# EDMP Template

---

- Intended to promote consistency and completeness
- A common format improves engagement and familiarity
- Living Document

# EDMP Template

---

- Introduction
- Personnel
- Systems
- Procedures
- Data Security
- Retention and Distribution

# Valid Values Management

---

## Best Management Practices in an Environmental Data Management System

Why valid values are critical to successful data  
management systems

# Why Valid Values Management is Critical

---

- Controls the input of new data
  - Not just limited to text values from look up lists
  - Includes meeting data type, field length and numerical range requirements to be considered *valid*.

# Topics Covered

---

- Common Terminology and Types
- Authoritative Resources
- Management Processes
- Additional Considerations for Managing Valid Values



# Valid Values Terminology

---

## ● Valid Value Synonyms Defined

- Allowed Values
- Controlled Lists
- Domain Values
- Enumerated Lists
- Lookup Values and Lookup Tables
- Lookup Lists
- Reference Values and Reference Tables

# Common Valid Value Types

---

- ◉ Analytical Chemistry
- ◉ Biological Assessments
- ◉ Geological
- ◉ Hydrogeological
- ◉ Locational
- ◉ Sample
- ◉ Consequential vs. Non-Consequential

# Authoritative Resources

---

- Federal Scientific Agencies
  - DOD, DOE, EPA, FWS, NIH, NOAA, USGS, etc.
- State Regulatory Agencies
- Institutes and Professional Organizations

# Authoritative Resources

---

- Federal Scientific Agencies
  - DOD, DOE, EPA, FWS, NIH, NOAA, USGS, etc.
- State Regulatory Agencies
- Institutes and Professional Organizations
  - TNI – The NELAC Institute
    - CSDP – Consensus Standards Development Program
    - NEFAP – National Environmental Field Activities Program
    - NELAP – National Environmental Laboratory Accreditation Program
    - NEPTP – National Environmental Proficiency Testing Program
    - SSASP – Stationary Source Audit Sample Program

# Valid Values Management Process

---

- Roles and Responsibilities
  - EDMS Stakeholders
  - EDMS Data Management Team
  - Subject Matter Experts
  - Authoritative Resources

# Valid Values Management Process

---

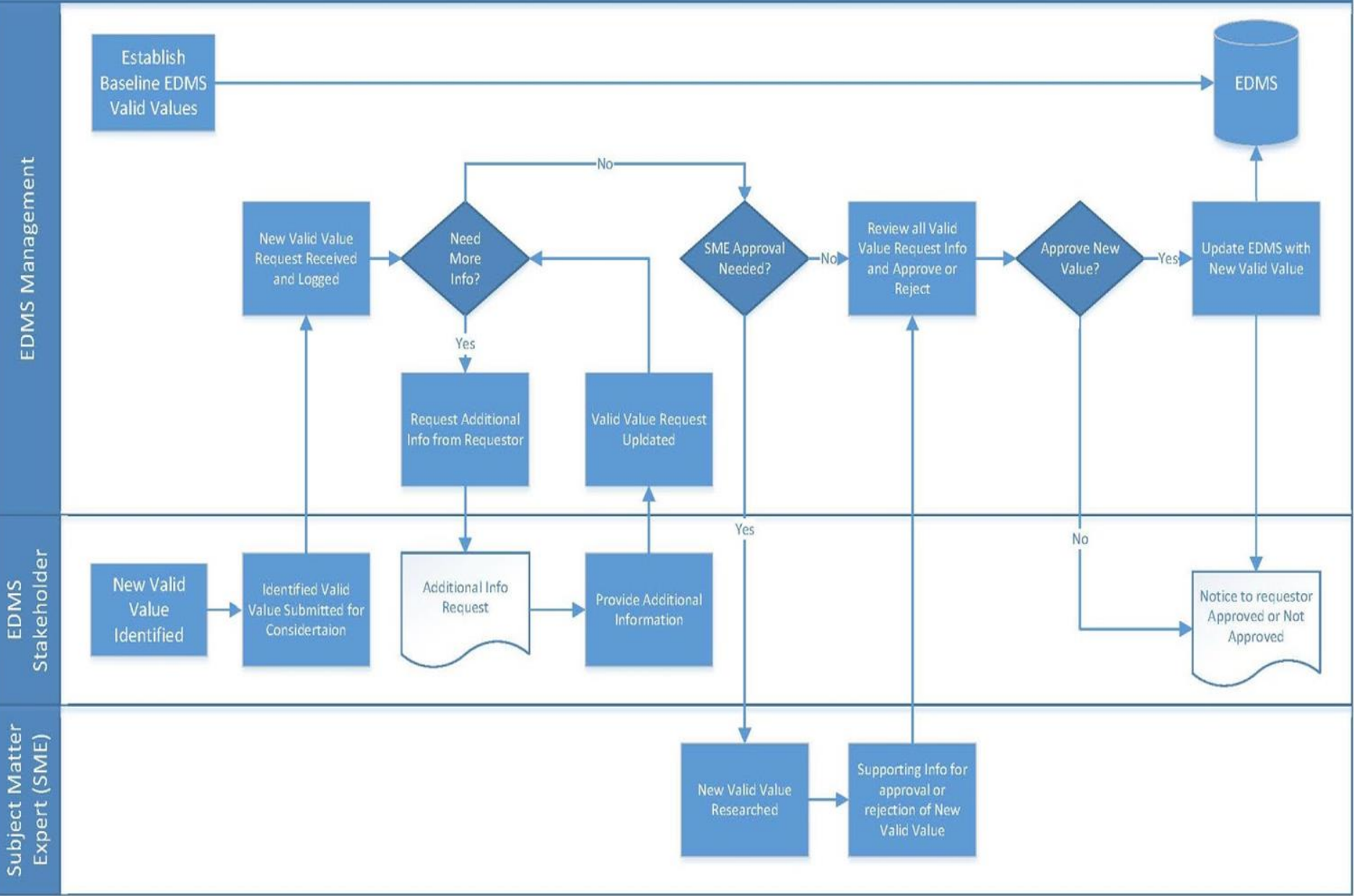
## ● Sample Valid Values

### Management Process Steps

- Process for requesting new Valid Values
- Process for approving valid values
- Standards for the creation of new valid values



# ICEDM Valid Value Management Process Example



# Additional Considerations for Managing Valid Values

---

- Upstream and Downstream Effects
- Methods for Managing Valid Values  
Synonyms
- Importance of Communicating Valid Values and Providing Valid Values Help

# **Analytical / Laboratory EDDs**

---

## **Benefits of Using EDDs**

**Defining the scope of the minimum  
requirements for an EDD**

# What are EDDs/SEDDs?

---

- Electronic Data Deliverable (EDD)
  - Computer readable
  - Pre-defined structured format
  - Popular formats include comma/tab delimited or ASCII files that can be opened as spreadsheets or database tables.

# What are EDDs/SEDDs?

---

- Staged Electronic Data Deliverable (SEDD)
  - Hierarchical Structure as part of the defined guidelines to denote project structure.
  - Usually in XML format which can also be read into database tables
  - Non-proprietary

# Benefits of Using EDDs

---

## ● Consistency

- Clearly defined reporting requirements

## ● Time Savings

- Structured format for creation of tables, figures, and reports.

## ● Data Quality

## ● Robust Program



# Current EDD Industry

---

- High Variability in EDD Formats
- Various Data Management Systems and Processes
- Labs Producing > 250 EDD Formats
- Efforts Focused on Mapping Data Between Various Formats

# Determining Minimum EDD Requirements

---

- Project Requirements
- Data Management Needs
- Data User Skill Level
- Value Added to Project
- Handling of Field Information

# Determining Minimum EDD Requirements

---

## ● Consider

- Regulatory Requirements
- Project Data Quality Objectives
  - QAPP Specifics and Data Validation Requirements
- Laboratory Reporting Capabilities
- Environmental Data Management Systems
- Stakeholder Requirements

# Determining Minimum EDD Requirements

---

## ● Consider

- Laboratory Capabilities and LIMS
- Valid Value Requirements
- Data Quality Objectives
- EDD Structure Requirements

# Minimum Requirements

---

## ● Basic EDD

- Core laboratory data required for simple data management

## ● Standard EDD

- Comprehensive EDD

## ● Advanced EDD

- 3<sup>rd</sup> Party Validation
- Rigorous regulatory requirements

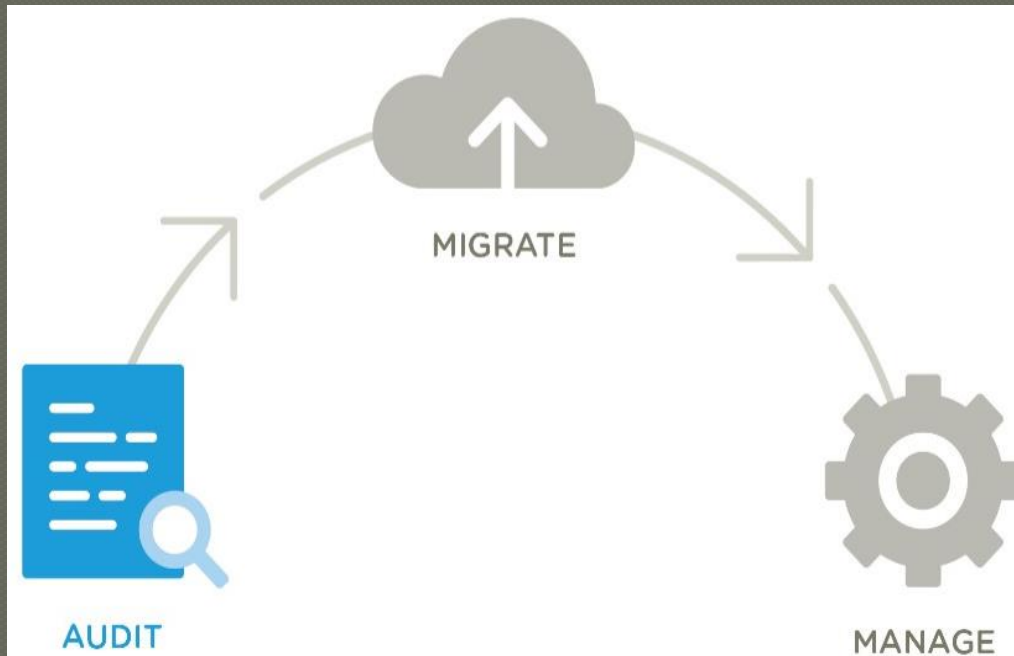
# Historical Data Migration

---

## Environmental Data Quality Audit: Foundation and Framework

“Know what you’ve got so you know where you’re going”

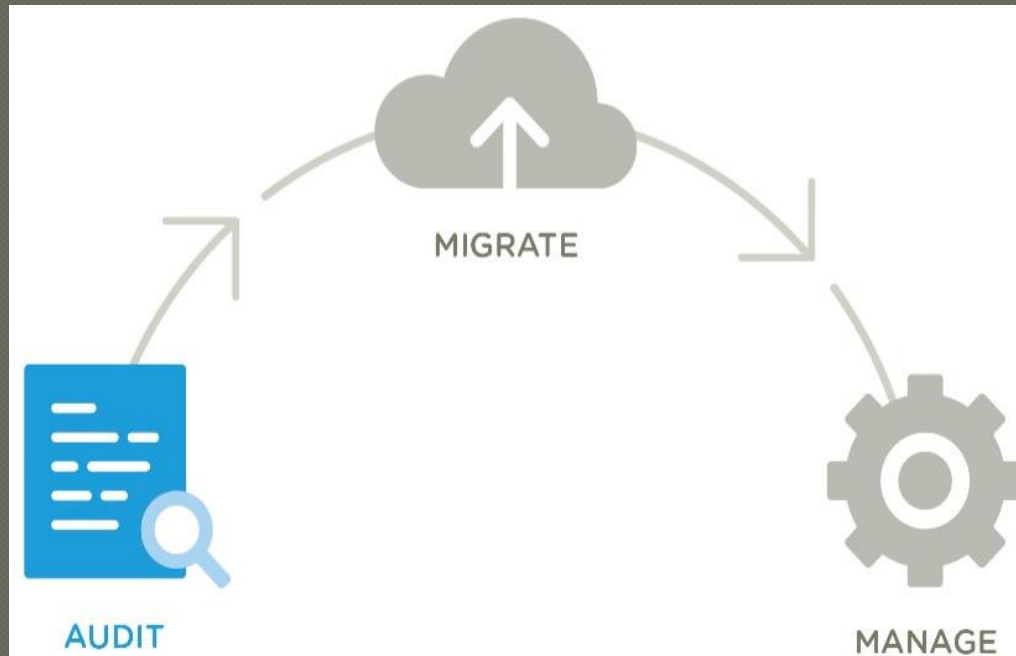
# What is a Data Audit?



- ◉ Inventory
- ◉ Evaluation



# What is a Data Audit?



## Why Audit Environmental Data?

- Reducing Risk
- Maximizing Quality

- Inventory
- Evaluation

# Data Inventory

---

- Goal: capture all distinct sources of information for evaluation
  - Make a metadata catalog

# Data Inventory

---

## ● Data Categories: organization method to manage expected data

- **Location** (coordinates, datum, well construction, lithology)
- **Analytical** (laboratory, validation)
- **Field** (depth to water, well stabilization, sample collection)
- **Continuous or High Density** (transducers, pumping meters, data loggers)
- **Investigative** (GIS, cross-sections, CSMs, DMP, QAPP)

# Example Data Inventory

	Laboratory EDDs		Project Database		Semi-Annual GW reports		Remedial Investigation Report		
File Attributes									
# of Files	15		1		13		1		
File Name(s)	ProjectA-[SDG#].txt		EDMS.mdb		SemiAnnual_2Q12.pdf,		RI_Contaminated Area.pdf		
File Location	ProjectFiles/DataFrom/ConsultantA/Data				ProjectFiles/DataFrom/ConsultantA/Reports				
File Type	Text Files		Access Database		PDF Report (selectable), with		PDF Report (scanned), with tables/maps		
File Format	ESBasic EDD		Custom Database Schema		PDF (selectable)		PDF (scanned copy)		
Source / Author	Lab Delta		Consultant Alpha (incl. data from multiple sources)		Consultant Alpha		Consultant Omega		
File Date(s)	2/2015 - 10/2016		October-16		4/2012 - 12/2016		March-10		
Data Provider	Consultant Alpha		Consultant Alpha		Consultant Alpha		Consultant Alpha		
Notes	Soil/GW/Air Investigation Data		Soil Inv & GW Monitoring data		SemiAnnual_2Q12.pdf, etc.		dated 3/3/10		
Location Data									
Coordinate System	None		State Plane		No		State Plane		
Spatial Datum Information			Horizontal	NAD27	No		Horizontal	NAD27	
			Vertical	Unknown			Vertical	NAD83	
Point Location Data			SBs, MWs - limited, some coordinates		MWs - very limited info		SBs, MWs - survey info Air data - limited		
Lithology / Boring Logs			Yes		No		Yes - appendix		
Well Construction	Yes - limited		No		Yes - appendix				
Analytical Data									
Attributes	Start	February-15	Start	July-99	Start	February-12	Start	July-99	
	End	October-16	End	October-16	End	October-16	End	July-09	
	Media	Soil/GW/Air	Media	Soil/GW	Media	GW	Media	Soil/GW/Air	
Notes			Pre-2005 - Consultant Zeta 2005-12 - Consultant Omega 2012-16 - Consultant Alpha		Summary cross-tab tables, all likely included in project database.		summary cross-tab tables. Pre-2005 data - collected by earlier consultants,		
Data Validation Results	None		Some yes, some difficult to tell		Yes - includes qualifier definitions		pre-2005 Unknown - does not specify on historical data  post-2005 assumed yes (this consultant contracted with 3rd party validator)		
Field Data									
Water Levels	None		None		Yes - table		Yes - table		
Field Measurements					Yes - scanned		None		
Well Stabilization / Purge					Yes - scanned		None		

# Data Evaluation

---

- Goal: ask questions about each dataset to find out the quality, reliability, and depth; get a sense of how it fits into the dataset as a whole.

# Data Evaluation

---

## ● Establish Evaluation Criteria

- **Data Comprehensiveness** (no data vs unusable/incorrect vs complete)
- **Data File Format** (hand-written, scanned PDFs vs electronic, well-structured)
- **Temporal Completeness** (date gaps vs complete project lifespan)
- **Data Normalization / Valid Values** (inconsistent nomenclature and attributes vs normalized with unique valid values)
- **Source Tracking / Metadata** (no source vs robust metadata)

# Data Evaluation

---

## ● Evaluation by Data Category

- **Location** (Consistent naming? Coordinates display on site?)
- **Analytical** (enforced data integrity? Duplicate data? Best result identified? Sources available? Comprehensive?)
- **Field** (Duplicate data? Transcription errors leading to unusual results?)
- **Continuous or High Density** (Is aggregation method known? Compare aggregation to source?)
- **Investigative** (Use to compare and truth other data)



## Example Data Evaluation: Analytical Data

	Laboratory EDDs		Project Database		Semi-Annual GW reports		Remedial Investigation Report	
File Attributes								
# of Files	15		1		13		1	
File Name(s)	ProjectA-[SDG#].txt		EDMS.mdb		SemiAnnual_2Q12.pdf,		RI_Contaminated Area.pdf	
File Location	ProjectFiles/DataFrom/ConsultantA/Data				ProjectFiles/DataFrom/ConsultantA/Reports			
File Type	Text Files		Access Database		PDF Report (selectable), with		PDF Report (scanned), with	
File Format	ESBasic EDD		Custom Database Schema		PDF (selectable)		PDF (scanned copy)	
Source / Author	Lab Delta		Consultant Alpha (incl. data from		Consultant Alpha		Consultant Omega	
File Date(s)	2/2015 - 10/2016		October-16		4/2012 - 12/2016		March-10	
Data Provider	Consultant Alpha		Consultant Alpha		Consultant Alpha		Consultant Alpha	
Notes	Soil/GW/Air Investigation Data		Soil Inv & GW Monitoring data		SemiAnnual_2Q12.pdf, etc		dated 3/3/10	
Analytical Data								
Attributes	Start	February-15	Start	July-99	Start	February-12	Start	July-99
	End	October-16	End	October-16	End	October-16	End	July-09
	Media	Soil/GW/Air	Media	Soil/GW	Media	GW	Media	Soil/GW/Air
Notes	Soil/GW field results present in project database.  Includes air results and lab QC not present in database		Separate detailed evaluation saved in document.  Pre-2005 - Consultant Zeta; minimal metadata 2005-12 - Consultant Omega; limited metadata 2012-16 - Consultant Alpha  Lab QC - none except MS/MSD		Summary cross-tab tables. No unique analytical results - all exists in project database		summary cross-tab tables. Pre-2005 data - collected by earlier consultants,  Only source of air sample results prior to 2015.	
Data Validation Results	None		Pre-2005 data - unknown, can't tell from records. 2005-12 - assumed validated, only final qualifier populated. 2012-16 - validated, with validation metadata (who, when, val level,etc.)		Yes - reference for qualifier definitions		pre-2005 Unknown - does not specify on historical data  post-2005 assumed yes (this consultant contracted with 3rd party validator)	
Review Criteria								
Data Comprehensiveness	good		ok		minimal		minimal	
Data File Format	good, but preliminary		good		ok - PDF data is selectable		poor - need OCR and review	
Temporal Completeness	2015-2016 complete (1999-2016 full timeframe)		ok - missing all air results and QC		ok for 2012-16 GW timeframe, but duplicate to project database		1999-2009 complete (1999-2016 full timeframe)	
Data Normalization / Valid Values	very good (inconsistent with other lab EDDs)		ok to poor - valid values used, but contains duplications. Same info stored in multiple places, some of which is different.		very good		very good	
Source Tracking / Metadata	very good - lab reports available		ok - laboratory and SDG info listed, but no task associations or source file names		very good		good - some info in the text on sources, labs, etc.	

# Data Audit Matrix:

		Review Criteria					Comments	
		Data Comprehensiveness	Data File Format	Temporal Completeness	Data Normalization / Valid Values	Source Tracking / Metadata		
Data Categories	Location Data	Coordinates						Missing pre-2010 coordinates
		Coordinate System Defined	N/A	N/A	N/A			Consistent
		Datums Defined	N/A	N/A	N/A			Consistent
		Lithologic Logs						Exist for wells installed in 2010 only
		Drilling Info						Limited electronic data
		Well construction logs						Limited electronic data
	Analytical Data							Best source is project database; older data unvalidated.
	Field Data	Well Stabilization						Complete detail since 2012, older data missing.
		Water Levels						Spreadsheet source only; questionable calculations
	Continuous or High Density Data	Transducers			N/A	N/A	N/A	Transducers measured in 2009, but no data provided other than a summary. Follow up with former consultant.
	Investigative Data	Site GIS Features						Building and parcel GIS features only
		Geologic Cross-Sections						PDFs only
		Site Model						Older Rockworks model

# White Paper Key Summary

---

- ◉ Data quality audits provide a foundation for a data migration
- ◉ Capture decision-making documentation
- ◉ Data inventory, evaluation, and matrix provide different presentations and depths of the audit
- ◉ Audits support good data stewardship throughout the project lifecycle

# ICEDM Best Management Practices Group

---

## ● Path Forward

- Finalize and publish white papers
- Begin development of next round of white papers
  - Data Management Plan
  - Analytical/Laboratory EDDs
  - Historical Data Migration
  - Valid Values Management
  - Mobile Data Collection

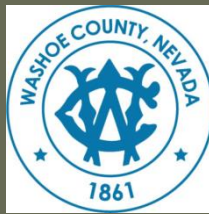
# 2017 Best Practices Round Table Discussion

---

- Field Teams: Data Collection
- Legacy/Historical Data
- Standardizing Field Forms
- Security
- Data Deliverable Formats
- Define Terms
- Training
- Communication
- Team Definition
- Collection Method
- Audits
- Change Management/Audit Trail
- Needs Assessment
- Data Management Team
- Define Terminologies
- Data Management Plan Elements
- Data Management Basics
- Data Visualization

- Analytical:
  - Best Result
  - Qualifiers
  - Aggregated/Derived Data
  - Detected Data
- Valid Values
- Ethics
- Workflow
  - Defining Data System's Purpose
  - Validator
  - Laboratory
  - Field Crews
- Public Data Availability
- Data Verification Levels
- Nomenclature Guidance
- Maintaining query parameters for reporting
- Qualifiers/Reason Code





## Q&A / Open Discussion

