ITRC has developed a series of fact sheets that summarizes the latest science, engineering, and technologies regarding environmental data management (EDM) best practices. This fact sheet describes:

- best practices for developing valid values to maintain consistency and reduce conflict or data loss when exchanging data with external systems
- best practices and considerations when changes to valid values are needed
- considerations for communication of valid values
- examples of the types of data fields that may require valid values

Additional information related to data exchange is provided in the fact sheet on Electronic Data Deliverables and Data Exchange, USGS Challenges with Secondary Use of Multisource Water Quality Monitoring Data Case Study, and Historical Data Migration Case Study: Filling Minnesota's Superfund Groundwater Data Accessibility Gap.

1 INTRODUCTION

In environmental data management (EDM), certain data fields have a limited number of acceptable values. Examples include whether a well is dry, which analytical method was used, what is the measurement unit of a value, or which projection or datum is used for a spatial coordinate.

Environmental data management systems (EDMSs) should have controls in place to ensure that values entered in restricted data fields are acceptable. The most common way to enforce controls is by developing and maintaining lists of accepted values, called valid values, allowable values, domain values, or reference values. Valid value lists are stored in reference tables, also known as lookup tables.

If controls aren't used, multiple instances of the same information can occur within a single data field. This data can be difficult to reconcile and can result in lost data or reduced data integrity. Additional effort is often required to recombine the data, using time-consuming techniques such as back-end database updates. Additionally, control of restricted data values provides clear, unambiguous, and consistent definitions of the data (for example, values for sampling method might include specific pump types if such granularity is desired). Clear, unambiguous, and consistent data definitions are especially useful when values might change over time. Examples include changes to scientific or common names of biological species or analytical methods where new versions are given a unique code or name that clearly defines the version used to quantify each result.

Example of Redundant Versions Without Control of Value

- monitoring well
- well-monitorign (Note: misspelling)
- well-monitoring
- observation well
- wells-monitoring

2 DEVELOPMENT OF NEW VALID VALUES

Development of new valid values typically occurs when an EDMS is first established and continues throughout its active life. Initially, many valid values in multiple data fields will be needed. After the initial setup, new valid values get added as the need arises. The following are points to consider when developing valid values for an EDMS.

2.1 Involve Subject Matter Experts

Subject matter experts or other staff with appropriate knowledge, including environmental chemists, scientists, or GIS staff, can help improve the overall quality and integrity of valid values.

2.2 Check Authoritative Resources and Adopt Accepted Valid Values

New valid values usually don't need to be generated from scratch. Adopt accepted valid values from authoritative resources if possible. Authoritative resources to check for existing valid values include non-EDMS-specific sources (for example, the U.S. EPA Substance Registry Services for names and CAS Registry Numbers to identify chemicals) or published reference tables from established EDMSs. See the Resources List for more examples.

The Minnesota Pollution Control Agency (MPCA) recently reviewed valid values from several other EDMSs when developing their Minnesota Groundwater Contamination Atlas. See the Historical Data Migration Case Study: Filling Minnesota's

Superfund Groundwater Data Accessibility Gap for details.

Note: When using valid values developed by other organizations, be aware that there could be errors in the definitions or that definitions might have changed over time. Valid values codes, especially, might also be unclear or dated. Check multiple sources to confirm the correct or best value for your EDMS.

2.3 Consider Sharing Valid Values with Other EDMSs

If data will be shared with or submitted to another EDMS, use of shared valid values can make data exchange a smoother process and reduce the risk of data loss due to incompatible valid value definitions. Keep in mind that using shared valid values might not always be practical or possible.

When an EDMS exchanges data with several other EDMSs, the likelihood increases that valid values and reference tables will differ among them. There may be no one set of valid values for any particular data field that will match all of the others. It might be necessary to remap certain valid values or use synonyms. Remapping involves making a crosswalk between valid values in one EDMS to valid values in another EDMS. This practice is essential for valid values that have different codes, but the same meaning or description. For example, a value of “air/climate” in one EDMS is equal to a value of “atmosphere” in another, based on the description.

2.4 Avoid Redundant Valid Values

More than one valid value with the same meaning or definition can lead to confusion and lost data during data reporting. For example, if both “W” and “Water” are used to represent a matrix or media type of “water,” a user searching for data might select only one of the values and therefore not return all pertinent data. Consider using synonyms if your EDMS accommodates them, or remap to a single reference value for each definition when multiple values are needed for different applications or for exchange with different entities.

When adding a new valid value to an existing reference table, make sure that a synonymous value isn’t already present. For example, “PCB-012,” “PCB-12,” “1,1′-biphenyl, 3,4-dichloro-,” and “3,4-dichlorobiphenyl” are names for the same chemical and therefore they share CAS Registry Number 2974-92-7. It is important to include only one of these names in your EDMS, along with the CAS Number. If your EDMS accommodates synonyms, synonymous names can be included. Synonyms for fields such as chemical name are useful for helping users find and choose the correct valid value, facilitating data exchange, and reporting results.

2.5 Develop Naming Conventions

Naming conventions keep current and future valid values consistent. Naming conventions are systems or rules that define the structure and source of the values and the process of developing new values. In most cases, established naming conventions make the addition of new valid values to existing reference tables a relatively simple matter.

Naming conventions may rely on authoritative sources, such as the Unified Soil Classification System for soil types, or may define the structure of the valid values, such as versioning within the code. For example, if an EDMS has existing analytical methods “SW8270” and “SW8270C,” the established naming convention indicates that a new value should be “SW8270D” and not “EPA8270 D.”

2.6 Establish Meaningful Valid Value Codes

Some valid values use abbreviations or codes (for example, data qualifier such as “J”), while others don’t need to (for example, chemical name such as “3,4-dichlorobiphenyl”). When codes are needed, create meaningful ones that can be understood without needing to look at a definition list. This helps streamline data management and reduce errors when using valid values.

Real world examples of ambiguous valid value codes include meaningless numbers like “2” (a horizontal datum code for “North American Datum of 1983”) or very short codes like “WLG” (a data qualifier for “Nearby wells flowing during measurement”). These codes are difficult, if not impossible, to understand without looking at the definition. A data entry error in one character can inadvertently change the value to an entirely different and incorrect code and meaning without a ready way to recognize or correct it. These types of ambiguous codes are often relics of older EDMSs, established when database storage space was limited.

Short codes are appropriate when they are common knowledge codes. Examples of common knowledge codes include reporting limit abbreviations like “MRL” (method reporting limit) and coordinate datum abbreviations, like “NAVD88” (North American Vertical Datum of 1988). Even with common knowledge codes, include a good definition.

Note: Be mindful of valid value code case. While most EDMSs are not case sensitive, many business analysis and GIS software packages are case sensitive. Variable case for a valid value can result in it being interpreted as two or more separate valid values by software that is coded to be case sensitive.

2.7 Provide Meaningful Valid Value Definitions

Valid value reference tables should provide a clear definition for each value. This helps the data management team locate gaps or overlapping definitions during development of the valid values. This also helps users select the correct value for data submission or remapping. Without clear definitions to assist in selecting or remapping valid values, data can be lost or become unusable during exchange or migration between EDMSs or other sources.

For example, lithologic descriptions (for example, mudstone or claystone) or monitoring location types (for example, ocean or marine) may differ between EDMSs, but with clear definitions, users can determine how to best choose or remap values with minimal information loss. Another example is data qualifier codes, which are often organization-specific because there is no universal standard. Data providers or labs use their own unique set of codes for specific qualification notes. Additionally, some are pre-validation lab flags while others are end-user, post-validation data qualifiers. Without clear definitions for remapping, the qualifier codes could be assigned completely different meanings when the data are shared with another EDMS.

Key points in providing meaningful valid value definitions:

- Valid values that are coded or abbreviated need to have the name or meaning spelled out in definitions (for example, data qualifier “J” is “Analyte was positively identified and the reported result is an estimate”).
- Valid values that some users may find obvious may be unfamiliar to other users or otherwise confused without a clear definition. For example, there are numerous units of measure with specific applications that users may not be familiar with (for example, “cSt” for “centistokes”). Even familiar units may have multiple options (for example, “U.S. Survey foot” versus “international foot”).
- Where future changes to valid value definitions are possible (for example, taxonomic scientific names), a clear definition in the original valid value reference table can help to reduce confusion or error during remapping to the new value.
- Even common knowledge codes require good definitions.

2.8 Tackle Uncertainties

Even with established naming conventions, cases arise where there is uncertainty about adding new valid values to existing reference tables. In these cases, consult other EDMSs, authoritative sources, and subject matter experts. Note that even authoritative sources sometimes have errors. Advice from subject matter experts can help to develop resolution in these cases.

2.9 Considerations for Data Exchanges

All of the preceding information is even more important if an EDMS is a state or federal system that has many points of data exchange, both in and out, with other systems and entities. New valid values added to these EDMSs will affect many other EDMSs and might conflict with their valid values. An EDMS with many points of exchange and a wider range of data will need a more comprehensive set of valid values and will need to anticipate valid values needs in advance.

The fewer points of data exchange an EDMS has, the more flexibility there is in defining valid values. A small or localized EDMS with limited data exchange might not need as extensive a set of valid values, instead limiting them to the specific data managed.

3 CHANGING VALID VALUES

Over time, it's likely that valid values in an EDMS will change. That may be because of errors or changes in the mapping of the original valid value (such as chemical codes for emerging contaminants); underlying data (such as changes in taxonomy definitions); or external systems that the EDMS exchanges data with.

3.1 Consider Cascading Effects in Advance

As with the development of new valid values, EDMSs with many points of data exchange need to consider the cascading effects of changes to valid values. A change made in an EDMS with many points of data exchange may require external EDMSs to make changes for data exchange to continue. Managers of other systems that exchange data with those EDMSs may also have to consider changes or additional remapping. EDMSs with fewer points of data exchange may have more flexibility to make changes as needed, but must consider the valid values of the external EDMSs that they exchange data with.

3.2 Develop a Change Management Process

Develop a process to identify and propose valid value changes; determine new valid values; document changes; and communicate changes. The formality of this process depends on the EDMS size, complexity, and degree of interconnection with other EDMSs.

- **For a small EDMS** the process may involve emails or calls within the data management team; a team member researching the new value; simple documentation of the change; and communication of the change with team members.
- **For a more complex EDMS** the process may involve a formal, documented request; documented research of the new value, possibly involving subject matter experts; formal documentation of the change; and communication of the change to external parties that provide data to or retrieve data from the EDMS.

3.3 Review Changes

Changing valid values requires review similar to developing new valid values. Always review existing valid values to confirm that synonymous or overlapping valid values don't already exist. Synonymous values are values with the same definition. Overlapping values have definitions where it may be unclear which value applies in some circumstances. For example, both "surveyed" and "GPS" are methods for collecting coordinate data. "GPS" could mean a professional surveyor used specialized GPS equipment and benchmarks for high accuracy, or a field team used a commercially available GPS unit at a lower degree of accuracy. "Professional survey" and "GPS consumer unit," might be better choices. As always, make sure the valid value definitions are clear.

3.4 Consider Remapping rather than Changing Valid Values

Changing valid values in your EDMS for the purposes of data exchange might not always be practical or necessary. Remapping valid values is usually a more efficient strategy. Remapping converts existing valid values to the new valid values prior to exchanging data. Examples of cases where remapping is beneficial include a new EDMS that has different valid values, or an existing data exchange where the other EDMS changes valid values.

4 COMMUNICATION OF VALID VALUES

Communication of valid values is essential and should include both the values and the definitions associated with each data field. When valid values for an EDMS aren't available to systems that must exchange data with it, the risk of lost or poorly remapped data increases during data exchange. Good valid value definitions make sure that all parties exchanging data are using valid values consistently and correctly.

4.1 Methods of Communication

Communication might range from a publicly accessible website that provides valid value reference tables, with a listserv or other means to communicate changes, to a spreadsheet of valid value reference tables sent to a single party. The

International Conference on Environmental Data Management (ICEDM) white paper Valid Values Best Management Practices in an Environmental Data Management System provides descriptions of several ways to communicate valid values with parties submitting data to EDMSs, including incorporating valid values into an electronic data deliverable (EDD) form or documentation; linking to reference tables from EDD documentation; or providing a reference table export (ICEDM 2017).

4.2 Valid Value Access

Making valid value reference tables publicly available and downloadable is helpful to all users of large EDMSs with many points of data exchange, especially for systems providing data to that EDMS. The benefit even extends to systems that aren't providing data to that particular EDMS. Publicly available valid value reference tables can help other EDMSs streamline and align their valid values. They can also help users who search or download data from a public EDMS. Valid value reference tables help users know what categories of data are available.

5 EXAMPLE LIST OF DATA FIELDS USING VALID VALUES

Table 1 is a noncomprehensive list of data fields that often use valid values. Depending on the types of data managed in an EDMS, valid values for additional data fields might also be needed. This list is intended as an example or a starting point for development of reference tables specific to an EDMS.

Table 1. Valid value data fields with descriptions and examples

Data Field	Description/Examples
General	
Units of Measure	Units of measure associated with the value
Unit Conversion	Factor to convert between units
File Type	Standard file extensions and mime file type codes (.docx, .xlsx etc.)
User Name	Name of person who made the measurement, collected the sample, ran analysis, surveyed, etc.
Geographic Data	
Location State	State abbreviation (NY, BC, MI, etc.)
Location County	County Code or Name
Location Country	Country abbreviation (USA, CN, DE, etc.)
Coordinate System	Decimal Degrees, State Plane Coordinate System, Universal Transverse Mercator (UTM)
Coordinate Datum	For example, WGS 84, NAD83HARN or HARN
Coordinate Collection Method	A method for how the coordinates were collected
Elevation Datum	NAVD88, WGS84, etc.
Elevation Collection Method	A method for how the elevation was collected, such as GPS, professional survey, or digitized
Surveyor Company Code	A code representing the company that completed the site survey
Sample Information	
Sample Matrix or Media	Soil, water (groundwater, stormwater, etc.), air (ambient, indoor, etc.), tissue (animal, plant), etc.
Sample Collection Method, Sampling Equipment	Method or equipment used to collect a sample (for example, Bailer, KemmerBottle-PVC, Pump-GW-LowFlow)
Sample Preservation Method	Procedure or method used to preserve a sample
Sample Container Type	Type of container

Quality Control (QC) Sample Type, QC Codes	Designation of whether the same is a QC sample, and if so, what type (for example, normal or natural environmental samples, field duplicate, lab duplicate, etc.)
Sampling Company, Lab, or Contractor Code	For example, ACME Corp, Stark Industries, Wonka Inc.
Field Data	
Downhole Point Parameter	Measurements captured during field work (resistivity, gamma ray, soil electric conductivity, etc.)
Parameter Aliases	Alternate names for parameters
Parameter Type	High-level grouping of parameters
Geological and Drilling Data	
Geologic Units	Geologic formation for sample or boring
Soil Classification or Formation Type, Lithology	System used to describe soil and lithology classifications
Soil Classification or Formation Type, ASTM Codes	System used to describe soil and lithology classifications
Well Casing Material	Material used in a segment of the well
Annulus Material	Material used to fill a segment of the annulus
Drilling Method	Method of drilling
Laboratory Analyses	
Chemical or Parameter	Unique identifier for parameter (chemical or other result)
Speciation; Chemical Form (sample fraction)	Portion or component of the chemical measured in a sample
Analytical Method, Result Method	Procedure or method used to derive a result
Preparation Method	Procedure or method used to prepare a sample for measurement or analysis
Data Qualifiers	Used by laboratories and third-party data validators to verify, qualify, and validate the data
Data Validation Level, Data Review Status	Level to which data were validated by a third party or other qualified data validator
Result Statistical Basis	Method used to calculate derived results
Detection Limit, Detect, Detect2, LimitType, LimitType2	Type of detection limit (MDL, RL, PQL)
Laboratory Sample Matrix	The matrix of a sample in the lab
Result Status	Raw, provisional, final, etc.
Biological	
Taxonomic Name	Scientific or common name
Taxonomic Identifier, Taxonomic Level	Taxonomic serial number
Life Stage	The life stage of the subject organism
Gender	Gender of the subject organism
Tissue Type	Type of tissue that was analyzed
Habit	Position the organism occupies in a food chain
Voltinism	Number of broods or generations of the organism in a year

Cell Shape	Cell shape of phytoplankton organism

6 REFERENCES AND ACRONYMS

The references cited in this fact sheet, and the other ITRC EDM Best Practices fact sheets, are included in one combined list that is available on the ITRC web site. The combined acronyms list is also available on the ITRC web site.