Minimum Requirements for a Global Standardized EDD Structure for Environmental Laboratory Data

Introduction

This White Paper has been developed by a committee of volunteers from the ICEDM conference attendees, to provide best-practice guidance and information to define minimum requirements for a standardized Electronic Data Deliverable (EDD) for environmental laboratory data.

The paper introduces a scaled approach to determining minimum laboratory EDD structure requirements and standards in guiding the creation of analytical laboratory EDDs, with the goal of producing an EDD structure that makes data use more efficient, thus saving time and money. At the same time, by promoting a scalable structure, the intention is for the guidance to be flexible and useable by all industry groups to aid in improving consistency in process and product that can be used by laboratories and end data users. In addition, the paper promotes the benefits of using EDDs for data transfer and automation, along with decision trees to support defining minimum requirements and communications to achieve successful receipt of EDDs from the laboratory.

While the focus for this paper will be on identifying a basic content structure aligned to user's needs, it is anticipated that future ICEDM White Papers will be written with a focus on defining valid values, data-quality checks, and other data-management topics

It is intended that this White Paper remain generic with respect to any Environmental Data Management Software (EDMS) that an end user might use and would, therefore, be the best-practice guidance applicable for users who want to generate basic tables to integrate laboratory EDDs into complex databases.

What is an EDD?

An Electronic Data Deliverable, or EDD, is a digital computer-readable data file in which the data is stored in a predefined and structured format, making it a suitable medium for sharing, manipulating, and using data. Traditionally, an EDD is a comma- or tab-delimited file; however, it can be in any format — a Microsoft Excel worksheet, a Microsoft Access database, XML file, or an ASCII file. The simplest definition of an EDD is that it is a structured data file that places a particular data item in a particular column/order.

A Staged Electronic Data Deliverable, or SEDD, has a predefined structure that allows data submitters and requestors to vary the field of data being reported within defined guidelines, based on factors such as project scope/stage, data availability, and lab capabilities. The overall format of the EDD remains consistent but allows for scalability within the format to only report the data elements that are appropriately required.

In effect, an EDD or SEDD is simply a structured data file based on rows (records) and columns (fields) of data, a concept that is common to almost all industries and disciplines in which data is managed. However, it is interesting that what would seem to be a very generic term of '*Electronic Data Deliverable*' appears to have been specifically adopted by the Environmental Services industry, highlighted by a simple Google search of the term, where all of the top 50 results are related to the management of environmental data.

The desired purpose and requirements associated with a laboratory EDD can vary greatly dependent upon the goal and use of that laboratory EDD. It is important to understand and address the final use and requirement goals when making the decision as to what minimum level of EDD will be required of both the laboratory and the project. Some questions that should be answered are the following:

- Are there regulatory requirements associated with the final laboratory EDD format and

content?

- Will there be project specific quality assurance requirements that will dictate the need for particular content to be present?
- What are the laboratory reporting capabilities and requirements associated with the project and user needs?
- Will the laboratory EDD be uploaded to an Environmental Data Management System, and if so, what are the requirements of the software for particular content?
- How will the data format type affect the end user's requirements?

Benefits of using EDDs

It is commonly agreed upon that clearly defined reporting requirements are essential to providing consistent analytical data of known quality to the end user; therefore, analytical EDD use has become a much more common request from laboratory data users, and data users have a wide and growing variety of reasons for requesting analytical EDDs.

For some users, working with analytical EDDs is simply a function of creating a time savings in being provided with laboratory data in a structured format with which to create their own analytical tables, figures and reports for their needs. For others it is a critical element in the creation of a robust and complete data workflow for their project, environmental data-management system, and validation requirements. In these cases, the need for a streamlined and consistent laboratory EDD format can be critical in reducing data inaccuracies and variations, thereby helping to save on resource time, effort, and costs associated with correcting data quality, integrity and consistency issues in a given dataset.

Therefore, receiving analytical data from the laboratory in an EDD format has the potential to add value to the data end user's work product.

Current Industry Experience

Despite the clear benefits and wide use of laboratory EDDs within the Environmental Services industry, the current requirements for analytical laboratory EDDs vary greatly from project to project, with no recognized industry standard laboratory EDD structure or format, which leads to each software vendor, laboratory, consultant, regulator, and end user using different processes, methodologies, and needs in defining their own specific laboratory EDD requirements.

While it is recognized that the end user's needs for the data can vary greatly from ease of information delivery, to data quality or validation support, to required regulatory or litigation needs, the sheer number of different laboratory EDD structures and formats in use points to a historic lack of collaborative guidance and control by the industry.

It is not uncommon to see environmental laboratories advertise their abilities to be able to produce data in over 250 EDD formats. While this may be seen by the laboratory's marketing manager as a good selling point, the reality for the laboratory's operations team is that the number of EDD formats is quickly becoming unsustainable, and the efforts to map new custom-made formats alongside updating or creating variants of existing formats is significant. This extra effort is a contributing factor to a higher degree of inefficiency and data quality issues within the administrative and reporting functions of laboratories, because they struggle to keep track of what EDD requirements are needed for which client or project.

Defining scope of a Minimum requirements EDD

Minimum laboratory EDD requirements will vary depending on the end use of the data, but also on the maturity of data-management practices in different geographies. The ongoing developments of new environmental programs in emerging economies around the world may require fewer fields of data, and these initial environmental program developments may lend themselves to a simpler application of the electronic data. While in North America, and to a lesser extent Western Europe and Australia, where environmental programs are more mature and established, data is managed and analyzed for a much broader range of needs, with regulators demanding a much higher level of data validation and quality-assurance and -control.

For any project, the evaluation of the minimum laboratory EDD requirements should be aligned to ensure that they fulfill the wider project objectives, such that the careful and appropriate management of the project data adds value and efficiencies to the overall project scope. It is, therefore, necessary to identify and understand the scope of what the data end user will be doing with the laboratory data to determine what the minimum laboratory EDD requirements are needed per project. In addition, project requirements do evolve over time, so building in flexibility to scale up or down the minimum laboratory EDD requirements is also beneficial.

Given this acknowledgement that the minimum requirements will vary based on project requirements, this White Paper has been structured around defining a scalable methodology for determining the minimum requirements of a laboratory EDD. In addition, the scalable nature of the EDD means that the EDD structure being put forward could also be adopted as a single Staged EDD format, with each of the levels of minimum requirement identified acting as a 'stage' within the Staged EDD. The White Paper, therefore, identifies three levels of minimum requirements as follows:

1. Basic EDD – an EDD containing the core laboratory data required to support simple data-management and -analysis activities, often in the absence of an environmental data-management system (EDMS).
2. Standard EDD – an EDD containing laboratory data that allows a more comprehensive assessment of data to support more advanced querying, analysis and comparative reporting through an EDMS.
3. Advanced EDD – an EDD containing a full suite of laboratory data fields in which project data requirements need to meet more rigorous regulatory requirements or other third-party validations.

To determine which level of minimum laboratory EDD requirements are appropriate, a project's initial requirements should be evaluated with some key questions about the scope of data-management needs for the project and the resources and skills of the data end user, such as:

- Are results tables sufficient or is a complex relational database required?
- Does creating and managing laboratory data within a database add value to the project?
- Is a suitable EDMS available for use on the project?

It may be determined that the project may only require results information and will not be compiled in a database, and that a Basic EDD is appropriate for immediate needs. However, consideration must also be given to what will happen if the project evolves over time; therefore, evaluating criteria for today and for the future allows flexibility if it is determined later that a database will be created. In such instances, requesting the more comprehensive Standard EDD at the outset of the project from the laboratory would allow the potential for growth without additional future laboratory request and confusion.

If an EDMS is determined to be required, the minimum laboratory EDD requirements grow to allow a broader picture of the laboratory data being submitted and a more comprehensive data assessment. As the scale of such projects grows, it then becomes likely that there will be multiple project stakeholders looking to access, share and analyze the laboratory data, and so EDD considerations need to include how the data can be transferred to additional databases as seamlessly as possible. In this case, it is likely that the laboratory or the data end user would not want to produce or manage multiple laboratory EDD formats for the same data, so choosing the laboratory EDD that will satisfy the most requirements of all stakeholders with the least shortfalls would be advisable.

The minimum laboratory EDD recommendations in this White Paper are intentionally limited to the electronic transfer of environmental chemistry data from the laboratory to the end user. The recommendations are not meant to be applicable for documenting additional project information, similar but not limited to, geology, geospatial, and operation and maintenance data. Of particular note is the need to consider the field-sourced sample-related information that will be shared with the laboratory through the Chain of Custody (CoC) and how this information will be associated to the analytical results. From a laboratory perspective, CoC-based field information such as sample date/time, sampler, etc., can neither be verified nor validated by the laboratory, so the value of asking the laboratory to report this secondhand data back to the end user through the laboratory EDD has to be evaluated, especially when the end user should already have this information directly from the field engineers as a separate data feed.

Table 1 provides some further examples of factors that need to be considered when evaluating the minimum laboratory EDD requirements for a project:

Considerations	Basic EDD	Standard EDD	Advanced EDD
	Simple data management and analysis activities, often in the absence of an EDMS.	Comprehensive assessment of data to support more advanced querying, analysis and comparative reporting through an EDMS.	Project data requirements need to meet more rigorous regulatory requirements or other 3rd party validations
Use of EDD Deliverables			
Use by an individual for reporting purposes only	X	X	X
Submittal of EDDs to an EDMS		X	X
Sharing data with other Stakeholders/Regulators			X
Data Reduction and Reporting			
Basic Report Tables	X	X	X
Exceedance Crosstabular Tables		X	X
3rd Party EDD Submissions			X
Validation			
Lab Verification Only	X	X	X
Validation - Basic QC Summary (Stage 2A/2B)		X	X
Validation - Stage III and Stage IV			X
Data Analysis/Reporting Software			
Off-the-shelf EDMS		X	X
Excel	X		
GIS / CAD	X	X (IF LINKED VIA EDMS)	
Project			
Multiple Methods	X	X	X
Multiple Labs		X	X
Real Time Decision Making		X	X
Screening level data needs only	X	X	X
Requirements (ie QAPP, other)			X
Potential for Growth		X	X
Sampling Frequency			
One-time sampling project	X	X	X
Monthly sampling, multiple locations		X	X
Multiple frequencies defined by regulatory permit (e.g. Superfund sites)			X

Table 1 : Considerations when evaluating minimum laboratory EDD requirements

Laboratory Considerations

Pre-project communication with laboratories regarding minimum laboratory EDD requirements is essential to ensure that needs are met. It is important to consider the type of laboratory in which requests are made. For example, is it a mobile laboratory, municipal laboratory, or large contract laboratory? Each will have different capabilities to create formats and record all the relevant content to be included in the EDD, such as more robust Quality Control (QC) parameters in the more advanced

EDD standards.

Ultimately there is little point in asking a laboratory to produce an advanced EDD just for the sake of it when a large number of the fields are designated for items of data the laboratory may not record or collect within its Laboratory Information Management System (LIMS), and, therefore, is always left empty in the final EDD deliverable, especially when these items of data in question would provide no real added value to the wider project (e.g., lab instrument analyst's name).

As such, Laboratory Information Management System (LIMS) vendors are also an important component of successful EDD generation. Having a laboratory LIMS that can capture and output the necessary data elements needed to support multiple agency and project needs in a required EDD structure should be seen as being part of the minimum requirements. The ability of a LIMS to exchange a defined set of result data elements directly from laboratory instrument to EDD without the need for any manual double entry of data is an important step toward comparing data quality across time and laboratories.

Valid Values Considerations

Establishing valid values is an important aspect of data analysis and data sharing. Accurate database querying is dependent on the user's ability to identify like for like data parameters across multiple data deliverables over time or laboratories. While not part of the scope of this White Paper, the use of standard terminology in field headers and valid values lists when designing and populating a laboratory EDD is essential.

Of particular importance within laboratory EDDs is to establish robust valid value lists to cover:

- Lab Analytic Method references
- Analyte / Chemical references
- Data Qualifier references
- Unit references
- Lab Test Classification references
- Lab Result Type references
- Lab QA/QC Sample Type references
- Lab Matrix references

In many cases, the laboratories themselves will have adopted specific references that are being used within their own LIMS systems, and, where feasible, consideration should be given to adopting the valid values used by the laboratory. By adopting the existing laboratory valid values, it cuts down on the need for the laboratories to re-map their values and any assumptions that might be made in the re-mapping decision process. However, adopting the existing laboratory valid values is really only feasible when a project has a dedicated database, with laboratory data being supplied by a single laboratory for the duration of the project.

When a project or database has the potential to need to receive data from multiple laboratories, it is advisable that laboratory-related valid values are laboratory neutral. For example, if the list of laboratory analytic method references in the project database is aligned to the internal method codes used by laboratory A, it is not appropriate to ask laboratory B to map their codes to those of their competitor at laboratory A, especially if the codes are not intuitive in their nomenclature to match them to generic laboratory methods.

Given the broadness of the topic of Valid Values, a more in depth discussion of valid values has been addressed in an alternate ICEDM White Paper.

Data Quality Considerations

All laboratory EDDs, regardless of structure, should be checked for completeness and accuracy upon delivery from the laboratory. Laboratories producing EDDs have many methodologies and capabilities to produce the requested EDD formats, and understanding how the EDDs are produced by the laboratory and what checks have occurred prior to delivery is an important step to understanding the quality of the deliverable. Here are examples of items to consider:

- Are the laboratory EDD and the associated accredited PDF laboratory report made at the same point in time and exported directly from the same LIMS database?

- If a data quality issue is found, can the lab resolve it by adjusting the mapping structure within LIMS so the issue is fixed for all subsequent EDDs produced?
- Are the EDDs created by the laboratory manually, with data copied and pasted from one software application to another, or are they manually typed from hard-copy instrument readouts such that errors are more inconsistent, requiring a more thorough review of laboratory EDD deliverables?
- Has the laboratory performed any additional data quality checks on the EDD prior to submission?

Once the laboratory EDD is received by the end user, a minimum-level EDD requirement check is required to ensure accuracy. Completeness checks should include verification that all samples, methods, matrices and target analytes are reported as requested by the laboratory EDD end user. The checking routine should also take into consideration all the documentation specific to the sample data package, which may include the accredited laboratory final report, modified analysis requests, the Chain of Custody (COC), the Sample Delivery Group (SDG) narrative, and other applicable documents or project communications with the laboratory. When data is being reported in more than one data tab within the laboratory EDD file, it is also essential to verify that the data linking fields from one tab to another are accurate, and that they do not result in any orphan rows internal to the EDD. The laboratory EDD should also be free of duplicate rows of identical data.

When a laboratory EDD is to be incorporated into a larger project dataset, via an EDMS, further accuracy checks should include verifying that correct valid values were used and that the laboratory EDD data can be relationally linked to the wider dataset without resulting in any orphan rows. A good example of verifying that the laboratory EDD data can be rationally linked to the wider dataset would be to ensure that sample IDs provided in the laboratory EDD are matched to sample IDs provided in any EDD or data submission from the field.

Ultimately the extent to which a laboratory EDD is validated and verified with regard to data quality checking and the methodologies employed to validate and verify that the data will reflect the project scope. A highly regulated large project with multiple stakeholders will inevitably require a more comprehensive and detailed data checking routine, while data validation routines on smaller projects may be satisfied by cross tabulating the laboratory EDD to confirm the presence of all analytes and result values.

Therefore, the broad nature of laboratory EDD data-quality checking requirements is such that only the core data-quality considerations are presented here in this White Paper. It is anticipated that a future White Paper will be developed to cover this topic in more detail, covering issues such as appropriate checks for summed total results (e.g., Total PAHs), and validation methodologies for applying data qualifiers.

EDD Structure/Format/File Considerations

The physical structure, format and file type for a laboratory EDD will largely be determined by the requirements of an EDMS being used on the project, and, therefore, it is not the intention of this White Paper to recommend a specific structure, format or file type for a standard laboratory EDD, with focus instead on ensuring that any laboratory EDD structure and format used includes the data fields and elements outlined for minimum requirements.

However, some key considerations and best-management-practice advice to take into account when the end user has a degree of flexibility on the laboratory EDD structure and format are:

- Include field headers.
- Avoid project and location-based data unessential to laboratory reporting.
- Use Valid Values whenever possible.
- If a minimum data element is not applicable or required for a data group, leave the record null opposed to omitting the field.
- Avoid redundant information
- Minimize the number of separate files or data tabs per laboratory EDD.

It should also be noted that in the recommendations below of the data fields that should be included in a minimum requirements laboratory EDD, the specific naming of these data fields may be defined differently within any EDMS being used, even though the purpose of the data fields is the same. For example, one EDMS may use the terminology 'Lab Method ID' while another might use 'Analytic Method Code'. Likewise, many EDMS will also include additional minimum requirement data fields within their laboratory EDD to meet the needs of the specific software's data model and workflow, such as unique record IDs or derived data based flag fields, derived from data elements in other fields using a mathematical, logical, or other type of transformation. For example, detect status (Y/N) can be derived from the lab qualifier field while sample source (field vs. lab) can be derived as an attribute of sample type. In some instances, these derived data fields have been included in the minimum requirements when it has been deemed that such derived data adds value to the laboratory EDD.

Laboratory EDD file-naming conventions will also be determined largely by the requirements of any EDMS being used; however, it is recommended that the laboratory EDD filename at a minimum should include the laboratory sample delivery group. Standards for laboratory EDD file naming should be included as part of the overall project planning and included within the Data Management Plan, which is discussed further in an alternate ICEDM White Paper.

Data Fields for Minimum Requirement EDDs

Presented below are the data fields that have been determined as being of a minimum requirement for inclusion in a laboratory EDD to meet best-management practices. The data fields have been categorized in three levels of minimum requirements previously discussed:

1. Basic EDD – an EDD containing the core laboratory data required to support simple data management and analysis activities, often in the absence of an EDMS.
2. Standard EDD – an EDD containing laboratory data that allows a more comprehensive assessment of data to support more advanced querying, analysis and comparative reporting through an EDMS.
3. Advanced EDD – an EDD containing a full suite of laboratory data fields in which project data requirements need to meet more rigorous regulatory requirements or other third-party validations.

Basic EDD Minimum Requirements

The Basic EDD user would not have any particular data specifications or requirements passed down from their regulators or other project stakeholders but need electronic data for their internal use, such as data archiving or simple table and chart generation. Laboratory EDDs in this format would not be suitable for data validation or submission to most current regulatory databases or commercial EDMS. This level of laboratory EDD is only expected to contain field-sample data with no laboratory QC sample data.

Minimum Data Elements:

- Sample ID
- Sample Date/Time
- Sample Matrix
- Analysis Method
- Filtered (Total or Dissolved)
- Analyte Code/ID
- Analyte Name
- Result
- Result and All Limits Unit
- Laboratory Qualifier
- Reporting Limit
- Chain of Custody ID
- Laboratory Name/ID
- Laboratory Comments

Standard EDD Minimum Requirements

The Standard EDD user would be looking to use the laboratory EDD file primarily for data interchange purposes to upload the laboratory data into an EDMS. Their requirements will, therefore, be greater than those of the Basic EDD user because they will be looking to do more in-depth analysis and interpretation of the project data, including comparing data over time and between different laboratory sources. In addition, this laboratory EDD includes some basic details for identifying and comparing the

results from laboratory QA/QC samples, and some basic validation checks.

Minimum Data Elements:

- **All Data Fields identified as a minimum requirement for the Basic EDD**
- Laboratory SDG
- Sample Receipt Date/Time
- Lab Sample ID
- Sample Type (Field, Blank, Lab Matrix Spike, etc.)
- Parent Sample ID
- Analysis Date
- Analysis Type (Initial, Dilution, ReAnalysis, etc.)
- Preparation Method
- Preparation Date
- Sample Preservative
- Dilution Factor
- Result Basis (Wet vs. Dry)
- Analyte Type (Target, Surrogate, TIC, etc.)
- Detect Status Y/N
- Final Result Y/N
- Reportable Result Y/N
- Method Detection Limit
- Validation Qualifier
- Final Qualifier
- Validation Date
- Validator Name
- Validation Level
- Lab Prep Batch
- Lab Analysis Batch

Advanced EDD Minimum Requirements

The Advanced EDD user would consider the laboratory EDD to be a complete replica of the laboratory data package. It would include all laboratory quality assurance and quality control data and include relevant raw data so that re-calculations of results can be performed. This level of laboratory EDD would be required when the laboratory data is to be submitted to regulators or validated by other project stakeholders.

Minimum Data Elements:

- **All Data Fields identified as a minimum requirement for the Basic and Standard EDD**
- Lab Storage Batch
- Receipt Temperature
- Column Number
- Quantitation Limit
- Subsample Amount
- Percent Moisture
- Analyst
- Instrument ID
- Retention Time
- Lower Lab QC Limit
- Upper Lab QC Limit
- Relative Percent Difference
- Relative Percent Difference Limit

Conclusion

This White Paper has provided guidance on best management practices for defining suitable minimum content requirements for a laboratory EDD for any given environmental monitoring and sampling project. The guidance recognizes that the scope of projects varies widely and thus minimum laboratory EDD requirements must be scalable and flexible to meet project demands. This paper has presented three levels of minimum requirements - a Basic EDD, Standard EDD and Advanced EDD - with the suggested data fields that would be included at each level presented below in Table 2. The scalable nature of the EDD structure being proposed also makes it suitable for being adopted as a single SEDD format, as well.

In addition, this paper has provided guidance on how to determine which level of laboratory EDD would be required based on project scope, along with other considerations with regard to laboratory capabilities, valid values, data quality checking, and file structures and formats.

It is, therefore, envisaged that this guidance, along with other complimentary ICEDM White Papers will assist anybody faced with the need to manage environmental laboratory data in line with industry best practices.

EDD Field	Basic EDD	Standard EDD	Advanced EDD
Sample ID	X	X	X
Sample Date/Time	X	X	X
Sample Matrix	X	X	X
Analysis Method	X	X	X
Filtered (Total or Dissolved)	X	X	X
Analyte Code/ID	X	X	X
Analyte Name	X	X	X
Result	X	X	X
Result and All Limits Unit	X	X	X
Laboratory Qualifier	X	X	X
Reporting Limit	X	X	X
Chain of Custody ID	X	X	X
Laboratory Name/ID	X	X	X
Laboratory Comments	X	X	X
Laboratory SDG		X	X
Sample Receipt Date/Time		X	X
Lab Sample ID		X	X
Sample Type (Field, Blank, Lab Matrix Spike, etc)		X	X
Parent Sample ID		X	X
Analysis Date		X	X
Analysis Type (Initial, Dilution, ReAnalysis, etc)		X	X
Preparation Method		X	X
Preparation Date		X	X
Sample Preservative		X	X
Dilution Factor		X	X
Result Basis (Wet vs Dry)		X	X

EDD Field	Basic EDD	Standard EDD	Advanced EDD
Analyte Type (Target, Surrogate, TIC, etc)		X	X
Detect Status Y/N		X	X
Final Result Y/N		X	X
Reportable Result Y/N		X	X
Method Detection Limit		X	X
Validation Qualifier		X	X
Final Qualifier		X	X
Validation Date		X	X
Validator Name		X	X
Validation Level		X	X
Lab Analysis Batch		X	X
Lab Prep Batch		X	X
Lab Storage Batch			X
Receipt Temperature			X
Column Number			X
Quantitation Limit			X
Subsample Amount			X
Percent Moisture			X
Analyst			X
Instrument ID			X
Retention Time			X
Lower Lab QC Limit			X
Upper Lab QC Limit			X
Relative Percent Difference			X
Relative Percent Difference Limit			X

Table 2 : Data Field Contents for Minimum Requirements EDDs