SEDD SPECIFICATION
Draft Version 5.1

For the Staged Electronic Data Deliverable (SEDD)

August 2005

This page intentionally left blank.

Table of Contents

This page is intentionally left blank.

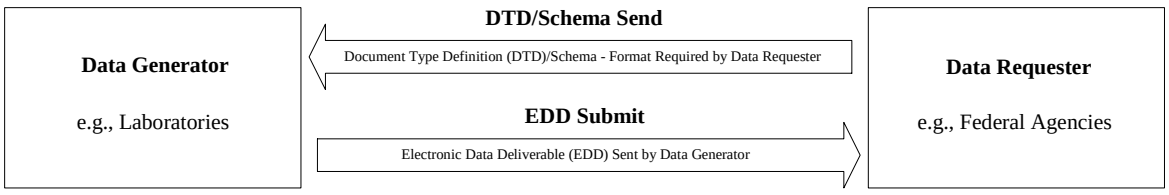
1.0 AN INTRODUCTION TO THE SEDD SPECIFICATION

What is the Purpose of the SEDD Specification?

The SEDD (Staged Electronic Data Deliverable) Specification is a specification for developing standardized electronic data deliverable formats for environmental analytical data. SEDD is designed to be Agency and Program neutral. An Electronic Data Deliverable (EDD) is comprised of actual electronic data that is delivered as a unit. The analytical data delivered by laboratories as EDDs to their data requesters includes sample identification, laboratory measurements, and laboratory Quality Control (QC) information.

The SEDD Specification provides the framework for developing the Document Type Definitions (DTDs) or Schemas and the resultant EDD formats by providing general specifications for the overall data structure of the EDD, while remaining flexible enough to be tailored for present and future individual Program or Agency needs. Figure 1 shows a general system for the transmittal of environmental analytical data from the data generator (the analytical laboratory) to the data requester (Federal or State Agencies, private firms, etc.). Data are generated at the analytical laboratory and sent to the data requester in the electronic format required by the data requester. These data are then checked by the data requester for format and content to ensure the data meets the data requester's contractual and technical requirements.

FIGURE 1. GENERAL SYSTEM FOR TRANSMISSION OF ENVIRONMENTAL ANALYTICAL DATA FROM THE DATA GENERATOR TO THE DATA REQUESTER



Each data requester must develop a guide based on the SEDD Specification and their individual technical requirements. This guide should include the DTD or eXtensible Markup Language (XML) Schema and a complete set of instructions for developing their specific individual EDD format. For the most common forms of environmental analysis (e.g., GC/MS, GC, ICP, ICP-MS), DTDs and Schemas have been developed for SEDD and must be used with no modification (see Section 1.5).

1.1 DEFINITIONS

The following terms are used throughout this document:

SEDD Specification
A specification for developing standardized electronic data deliverable formats for environmental analytical data. Under this specification, a common structure and dictionary are required.

Data Requester(s)
Individual(s) or organization(s) directly responsible for requesting analytical services and data from the analytical laboratory. For the purposes of this document, the data requester is assumed to be the person or organization responsible for developing the DTD or Schema. Examples of Data Requesters include Federal or State Agencies, private engineering or environmental firms, etc.

DTD (Document Type Definition)
The DTD provides the set of rules for developing the structure and data elements for specific XML EDD formats. These rules are established by the data requester and the SEDD structure.

Schema
The Schema (similar to but more powerful than the DTD) would give the set of rules for developing the structure and data elements for specific EDD formats along with the criteria for specifying the attributes of the data reported. These rules are established by the data requester and the SEDD structure.

Data Generator
Individual(s) or organization(s) directly responsible for generating data. For purposes of this document, the data generator is assumed to be the person or organization responsible for producing and transmitting the EDD or Schema. Examples of Data Generators include analytical, radio analytical, or field laboratories.

EDD (Electronic Data Deliverable)
An electronic file created by a data generator (usually the analytical laboratory) for transmitting and reporting analytical data.

XML (eXtensible Mark-up Language)
This is a standard devised by the World Wide Web Consortium (W3C) and provides a common approach to representing information over the Web. XML is a language for describing data. XML is not owned by any one vendor and thus remains an open standard. XML is text based, therefore it is processable using any platform. The SEDD Specification uses XML as its transmission standard.

Element

An element consists of a start tag, content, and an end tag. An element may contain other elements. The SEDD Data Element Dictionary (DED) contains the list of tags that developers may use. The SEDD Structure specifies relationships between certain elements from the DED.

DED (Data Element Dictionary)

The Data Element Dictionary contains the definitions of the defined data elements along with where in the structure they can be used.

1.2 RELATIONSHIP OF A DTD OR SCHEMA TO THE SEDD SPECIFICATION

The SEDD Specification provides the requirements for developing DTDs and Schemas and the resultant EDDs for reporting data to meet data requester's (e.g., a Federal Agency) needs. Final DTDs and examples of the resultant EDDs for major environmental methods are presented in separate files in Appendix C.

The SEDD Specification is flexible in that it can satisfy diverse requirements. The SEDD Specification uses a data model based on analytical activities as a starting point for requesting data deliverables. The SEDD Specification uses names for nodes and data elements to describe diverse types of laboratory activities. To take full advantage of the standardization available from the SEDD Specification, data requesters shall use these generalized names in their DTDs or Schemas. A Data Element Dictionary (DED) listing these generalized names is included in Appendix A. A DTD or Schema specifies which of the SEDD Specification nodes and data elements are required.

If, after careful review of the SEDD structure and the SEDD DED, a data requester is convinced that a particular critical requirement of their program is not addressed, specific data elements may be added to the data requester's DTD or Schema to address those needs. The data requester must be aware that implementation of specific data elements particular only to their Program reduce the ability of multiple Programs to share their data. Data requesters must present their specific data element needs to the SEDD Program Manager for possible future incorporation into the SEDD DED that is included within the SEDD Specification (see Section 2.1.2).

A DTD or Schema rarely requires all of the information available in the SEDD Specification. When necessary, a data requester can require additional information. However, all DTDs or Schemas and the resultant EDDs created using the SEDD Specification will use the same structure and DED.

This section, in conjunction with Sections 2, 3, and the SEDD DED, constitutes the Staged Electronic Data Deliverable (SEDD) Specification for developing the Document Type Definition (DTD) or Schema and the resultant EDD.

This SEDD Specification document is not a comprehensive specification for a specific DTD or Schema and EDD format. Specific DTDs or Schemas and the resultant EDDs must still be fully developed and defined by the data requester using the SEDD specification as a reference. In order to derive the greatest benefits from the XML technology, data requesters must utilize both the SEDD Specification structure and the SEDD Specification Data Element Dictionary.

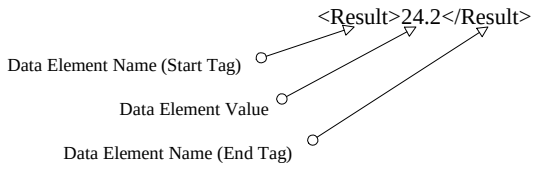
1.3 OVERVIEW OF AN EDD FILE CREATED USING A DTD OR SCHEMA BASED ON THE SEDD SPECIFICATION

A SEDD Specification-compatible EDD is a text file (with a .xml extension since XML is used as the transmission standard) that reports data in the form of "elements". There are two kinds of elements - data elements and nodes. Data elements contain data and have the following parts: a data element name (start tag); a data element value; and a data element name (end tag). The data element name is a descriptive label. The data element value is the data value associated with that particular data element.

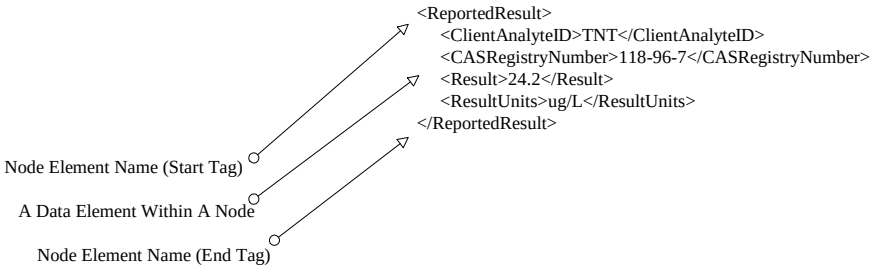
Data elements are contained within nodes. Nodes also consist of a start tag and an end tag, but instead of containing a data value between the start and end tag, nodes contain data elements between the tags. Figure 2 shows an example of data elements contained within a node. Nodes are arranged in a hierarchy. The level of a node in the hierarchy partially determines the node's meaning.

FIGURE 2. SEDD SPECIFICATION ELEMENTS

A. Typical Data Element and Value



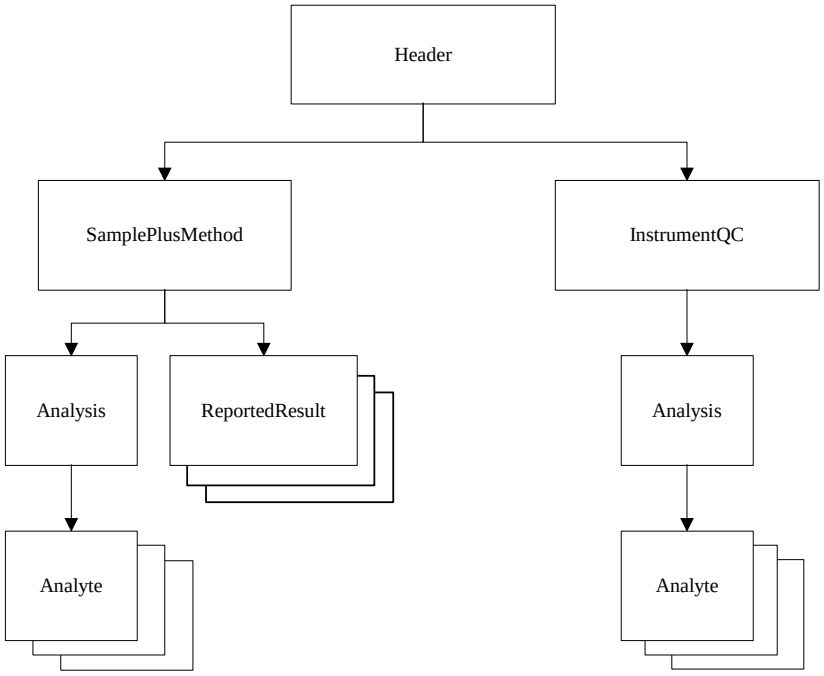
B. Data Elements Arranged Within a Node



1.4 THE SEDD SPECIFICATION HIERARCHY

The SEDD Specification hierarchy is based on a model of laboratory activities, the linkages between these activities, and the data these activities produce. To take a typical laboratory scenario, each sample analyzed by one method typically has several results that have to be reported (e.g., a volatile analysis by SW-846 Method 8260 would have many analytes whose concentrations are determined and reported). Information regarding the sample and method (sample identification, method used etc.) would be captured in the SamplePlusMethod node since this information would be the same for all volatile analytes being reported for the sample. The results of the analyses performed on the sample using this one method (e.g., the concentration of dichloromethane being 50 ug/L) would be captured in the ReportedResult node. There would be several ReportedResult nodes, one for each of the analytes whose results are being reported, with each of one of these nodes linked directly to the one SamplePlusMethod node (see Figure 3).

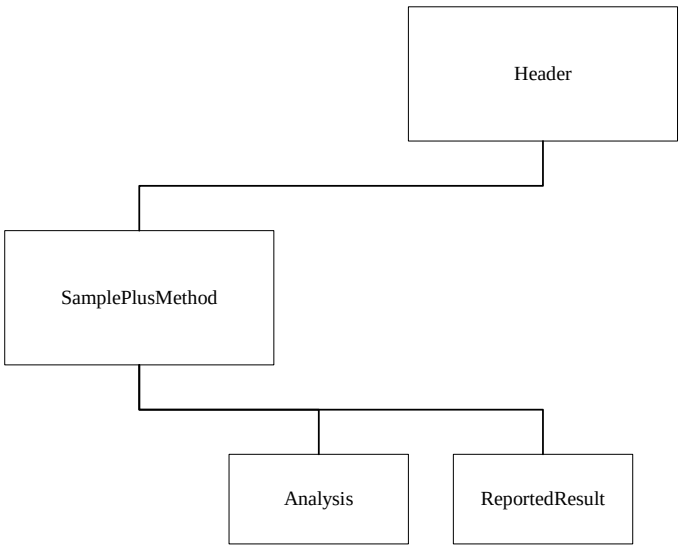
FIGURE 3. SIMPLIFIED SEDD SPECIFICATION HIERARCHY



To take full advantage of the standardization available from the SEDD Specification, data requesters shall use these structures in their DTDs or Schemas. To assist data requesters, the SEDD Specification structure can be implemented in the following three primary stages depending upon the level of detail the data requester needs in the EDD:

- **Stage 1** contains the minimum number of nodes and data elements to transmit "results-only" data. Only limited method QC data (or no QC data) would be reported in Stage 1. The Stage 1 structure is presented in Figure 4.

FIGURE 4. SEDD SPECIFICATION - STAGE 1



- **Stage 2** builds on Stage 1 by adding method (Stage 2a) and instrument (Stage 2b) QC data. The Stage 2a structure is presented in Figure 5 and the Stage 2b structure is presented in Figure 6.

FIGURE 5. SEDD SPECIFICATION - STAGE 2a

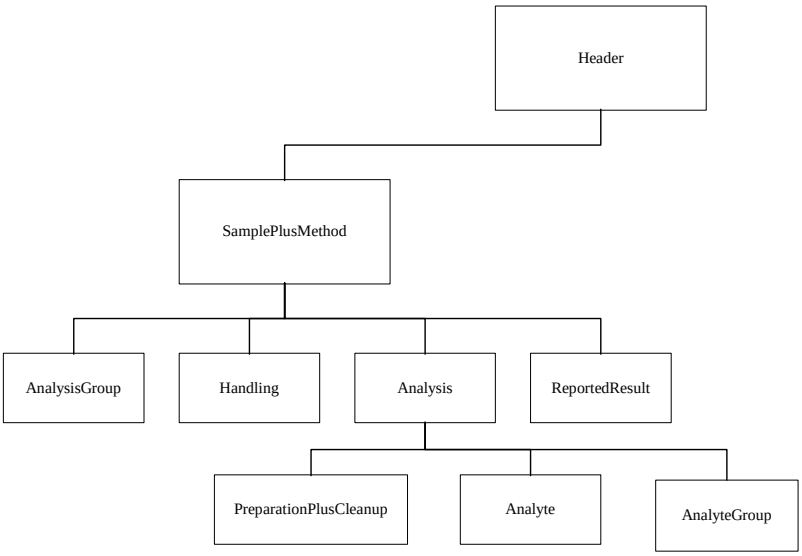
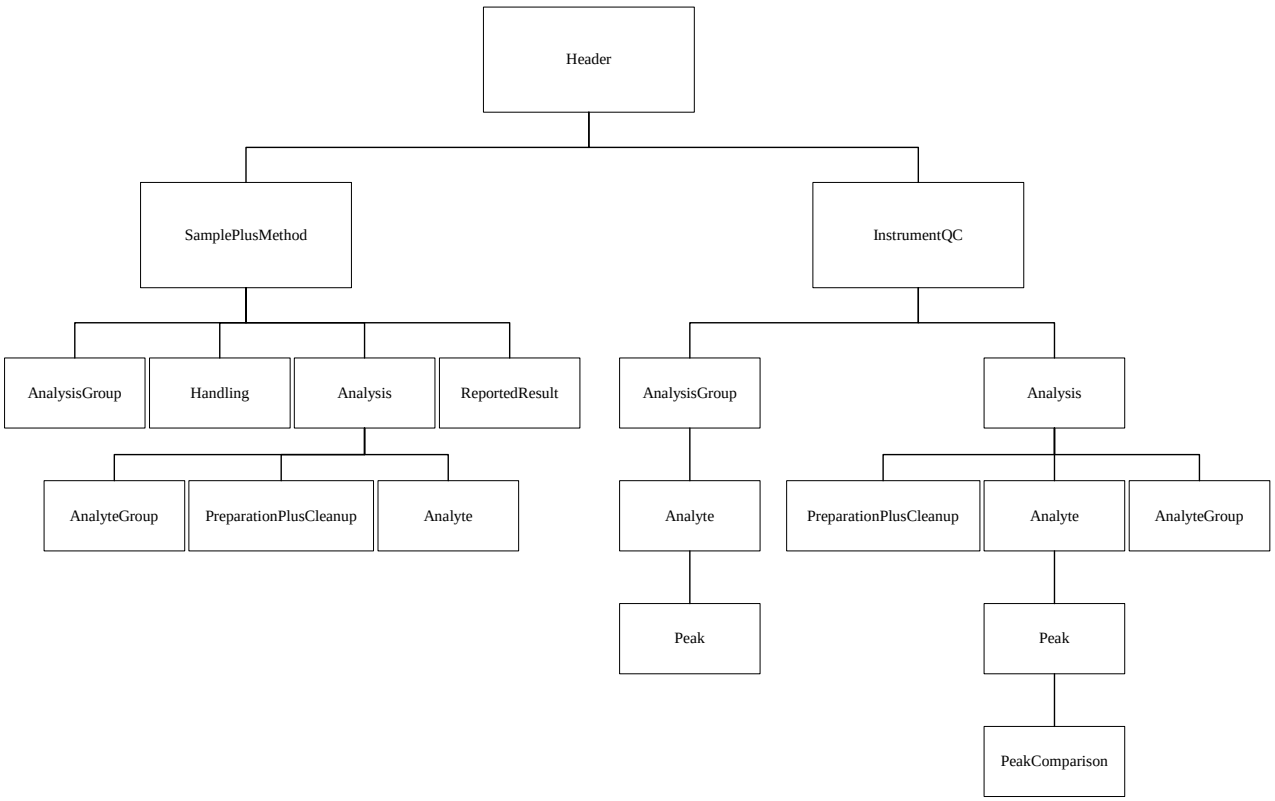
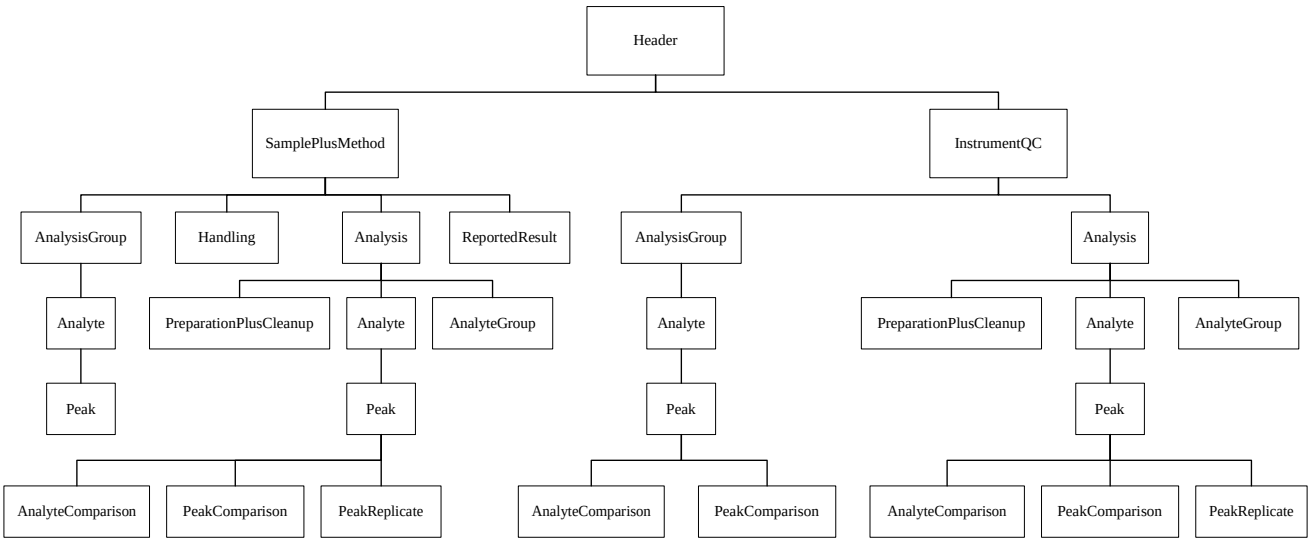


FIGURE 6. SEDD SPECIFICATION - STAGE 2b



- **Stage 3** builds on Stage 2 by adding additional measurement data to allow for the independent recalculation of the reported results. The Stage 3 structure is presented in Figure 7.

FIGURE 7. SEDD SPECIFICATION - STAGE 3



1.5 USE OF THE SEDD SPECIFICATION BY THE DATA REQUESTER

A. Common Environmental Analyses (GC, GC-MS, ICP, ICP-MS, CVAA, etc.)

For the most common forms of environmental analysis, program-neutral DTDs already exist and shall be used to ensure development of a uniform format for the transmission and mutual exchange of environmental analytical data. These DTDs, along with Valid Values for many data elements, are provided in Appendix C.

Program-specific requirements would be addressed in the instructions to the data generators.

B. Other Types of Analysis (e.g., Pharmaceutical, Biotechnology, Agriculture, Food and Beverage Testing, etc.)

Use of the SEDD Specification is not restricted to just environmental analysis, but can also be used for any type of chemical, radiological, or microbiological analysis data transfer. To create the DTD or Schema, the data requester must assess the current and future data needs of their program. Data requesters should evaluate which data elements (fields) they currently receive from their laboratories, either in an electronic deliverable or on a hardcopy form. Data requesters should also evaluate the level of assessment they need to apply to the data (e.g., check completeness only, or confirm calculated values). Based on this evaluation, data requesters then select the appropriate stage from the SEDD structure. This forms the basis for creating the node elements in the DTD or Schema using the required linkages.

Based on the specific data items required by the program, the data requester then adds the corresponding data elements from the SEDD DED to the appropriate node(s).

The DTD or Schema may also be used to define acceptable values for some of the elements. Data requesters may wish a certain value to be constant (fixed), or to come from a limited set of possible values. Developers should look under "Attributes" in an XML reference source for further details. The use of Schemas allow more flexibility in this area.

In addition to the DTD or Schema, the data requester must also specify the appropriate formats for each element. This would include details on identifying samples, methods, and projects, and specifying the minimum precision for measurements and results and the maximum length of any reported value. The following format considerations must be addressed:

- Define the level of detail required (e.g., Stage 1, Stage 2a, Stage 2b, Stage 3) provided in Section 1.

NOTE: It is strongly recommended that data requesters develop their data requirements for a Stage 3 DTD or Schema first, then pare their requirements down to a lower stage to ensure that the same DTD or Schema will work as data requester data needs grow.

- Use the SEDD Specification (i.e., use the selected stage which must include the required data elements, the linkages between the data elements, and the SEDD DED for defining all data elements).

- Define all valid values associated with appropriate data elements using the accompanying Valid Value list. Some elements may be restricted to a single valid value while others could have many valid values. For example the data element ResultUnits could have the following valid values associated with it: ug/L, ug/kg, mg/L, mg/kg, etc. A data requester may decide to only allow ug/L for Volatile Organics in a water matrix.

To assist the data generator in creating the resultant EDD, the data requester will also need to:

- Make the DTD or Schema and program specifications available to the data generator.

- Work with the data generator to clarify which specific data items already being supplied correspond to which specific SEDD data elements.

Sections 2 and 3 present the basic information to create an EDD based on the SEDD specification. Each data requester and data generator must become familiar with the guidelines and requirements given in these two sections. Once they become familiar with the guidelines, specific DTDs or Schemas and special instructions can be created by the data requester. The data generator would use this information to create the EDD.

Section 2 gives the basic rules for presentation of data in each data element, the hierarchy of the elements within the DTD or Schema and the resultant EDD, and the overall EDD file which must meet XML requirements.

Section 3 describes some of the key concepts underlying the SEDD Specification analytical model.

IMPORTANT NOTE: Sections 2 and 3 are not meant to be tutorials in XML technology. Only basic XML rules as applicable to the SEDD Specification are summarized. Both the data requester and data generator must be familiar with XML to correctly apply the SEDD Specification.

2.0 SEDD SPECIFICATION SYNTAX, REQUIRED DATA ELEMENTS, AND DATA ELEMENT VALUE FORMATS

This section provides the structural and data representation rules (i.e., syntax) for the SEDD Specification, along with the required data elements, and the format for each data element value. This information is broken into the following three subsections:

- Section 2.1 Syntax
- Section 2.2 Required Data Elements
- Section 2.3 Data Element Value Formats

Data can be transferred between a data generator and a data requester using XML technology only if the specifications regarding the data format have been worked out between the two parties prior to transmission of data (in the form of an XML file). The SEDD Specification provides the basis of the format for transmission of analytical data by clearly defining: the overall data structure; data elements; and relationships between the data elements.

NOTE: Most of the examples provided in Sections 2 and 3 pertain to Stages 2b and 3 of the SEDD Specification. At present, most laboratories deliver EDDs equivalent to Stage 2a of the SEDD Specification (sample results plus method QC data - see Figure 5). Pilot studies have demonstrated that most laboratories can implement EDD's based on a Stage 2a SEDD Specification format within a few months. Laboratories can then build on the Stage 2a EDDs to create and deliver Stage 2b (see Figure 6) or Stage 3 (see Figure 7) EDDs by simply adding additional nodes and data elements to the Stage 2a EDD (as data requestor needs change).

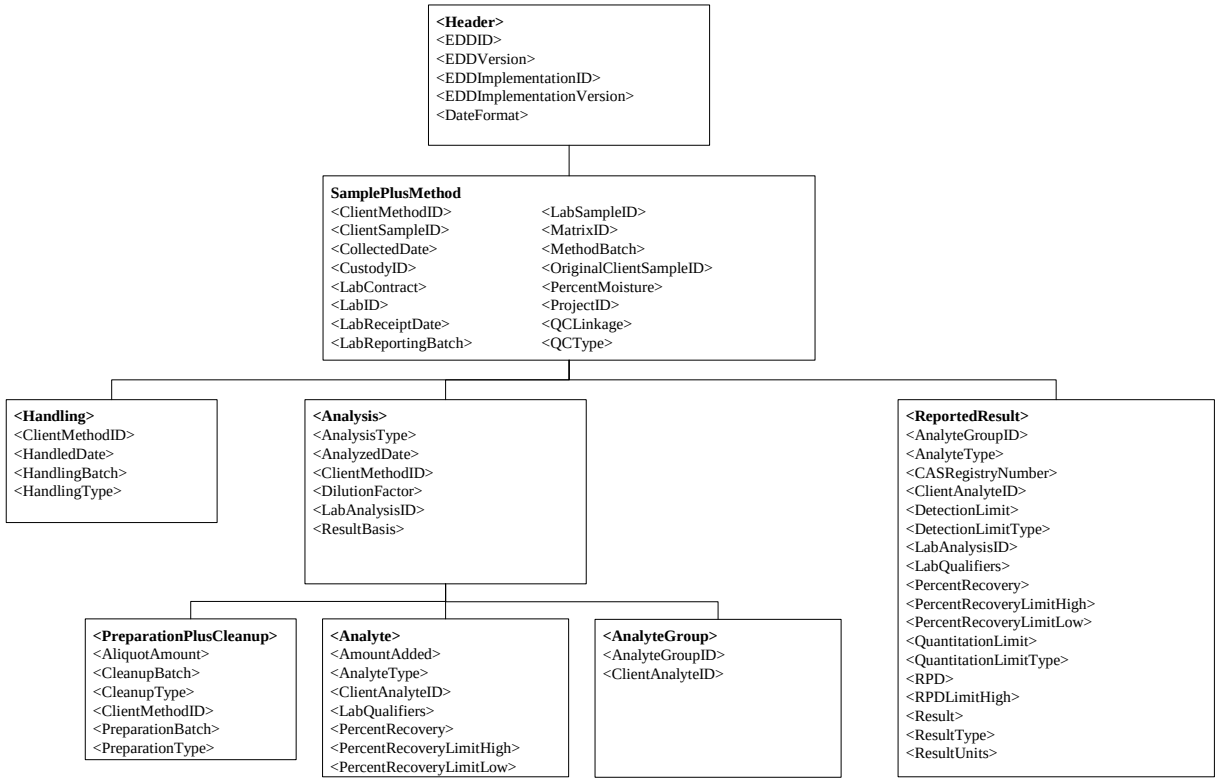
2.1 SYNTAX

Syntax is defined as the rules for representation of data and structure. This section describes the syntax for the following components of SEDD Specification-compatible EDDs, as well as the overall file syntax, as follows:

- Characters and Lines (see Section 2.1.1)
- Elements, Names, and Values (see Section 2.1.2)
- SEDD Specification Hierarchy (see Section 2.1.3)
- XML File Syntax (see Section 2.1.4)

Figure 8 is a graphical representation of a SEDD Specification-compatible EDD for a common organics method. This figure illustrates the major components discussed in the syntax section and provides a useful frame of reference for visualizing the relationship(s) between components. Figure 8 is not a complete example of the SEDD Specification syntax or data element list.

FIGURE 8. SEDD SPECIFICATION HIERARCHY AND DATA ELEMENTS FOR VOLATILE ORGANICS - STAGE 2a



2.1.1 Character and Line Syntax

Characters and Lines

- An EDD is a string of characters in a series of lines.

- The specific character set used would be specified in the opening XML declaration statement to call specific data linkage library modules/parsers within the operating system program of the computer.

NOTE: The encoding called UTF-8 (Unicode Transformation Format - 8-bit form) will be the default used for SEDD. The UTF-8 character set uses all 7-bit ASCII codes for values 0 through 127 and uses the ISO-8859-1 (Latin-1) codes for values 128 through 255. The use of other character codes is possible.

- There are six basic types of lines:

    1. **XML declaration line**: This is the first line in an XML document. Consists of optional leading spaces; a less than sign followed by a question mark followed by the letters xml; the xml declaration statements; a question mark followed by the greater than sign (e.g., <?xml version="1.0" ?>). The character set used by the EDD would also be specified in this same line (e.g., <?xml version=”1.0" encoding=”UTF-8"?>).

    2. **Document type declaration lines**: This is generally the second line in an XML document. Consists of optional leading spaces; a less than sign followed by an exclamation mark followed by the letters DOCTYPE; the DOCTYPE declaration statement; the greater than sign (e.g., <!DOCTYPE Header SYSTEM “General_3_1.dtd”>).

    3. **Blank lines**: Contain no characters. Blank lines can occur anywhere in an EDD. They can be used for visual formatting to aid browsers.

    4. **Comment lines**: Consist of optional leading spaces; a less than sign, the exclamation mark, two dashes; followed by the comment; followed by two dashes and greater than sign (e.g., <!-- This is a comment line. -->). A comment line does not need to include the word “comment”.

       Comment lines can occur anywhere in an EDD. Their content is not part of the information to be delivered by the EDD. They can be used to annotate an EDD for a browser and might be printed as part of a status report by a program that reads the EDD.

       NOTE: It is important to distinguish a <!-- comment line --> from actual comments made as part of the official EDD itself. Within an EDD, the Comment data element is used to convey information that may be important for the proper interpretation of the data presented (example: <Comment>This is a comment.</Comment).)

    5. **Data element lines**: Consist of optional leading spaces; a less than sign followed by a data element tag name followed by a greater than sign; the data element value; a less than sign followed by the forward slash sign followed by a data element tag name followed by the greater than sign (e.g., <DataElementTag>value</DataElementTag>). A data element usually delivers a single piece of information.

    6. **Nodes**: Consist of optional leading spaces; a less than sign followed by a node tag name followed by a greater than sign (e.g., <NodeTag>). A closing node tag name must also be used later in the EDD and consists of optional leading spaces; a less than sign, followed by a forward slash, followed by the same node tag name, followed by the greater than sign (e.g., </NodeTag>). A node contains other data elements.

- Optional leading spaces in lines are allowed so EDD data can be indented to improve readability.

2.1.2 Element Syntax: Data Elements, Names, and Values

Data Elements

- **Definitions for the SEDD Specification data elements are provided in Appendix A.**

- A data element tag consists of a start data element tag name followed by the data element value content followed by an end data element tag name (e.g., <Result>24.2</Result>). A data element usually delivers a single piece of information. Data elements are sometimes called fields.

- Spaces are not allowed unless they are an integral part of the data element value.

- The SEDD Specification has no rules forbidding the use of any data elements and data requesters cannot add any such rules. If data requesters receive more data elements than originally requested, the data requester should ignore these additional data elements.

Data Element Names

- A data element name is a string of alphanumeric (A-Z, a-z, 0-9) characters. A name identifies the data delivered by a data element.

- A name should be limited to a maximum of 30 characters.

- Character case is significant in names. The first letter of all words used in data element names must be capitalized. Example: AnalyzedDate would be acceptable while Analyzeddate would not.

- Data element names defined in SEDD are listed in Appendix A with a description of their usage.

- Implementation-defined names are strongly discouraged but allowed. They must begin with the word "New". It is recommended that an identifier for the implementation (i.e., the Agency responsible for the data) follow the word "New" (e.g., 'NewEPADataElement'). This will prevent conflicts across formats when implementation-specific names have been added. Useful implementation-defined names will be incorporated in later versions of SEDD under SEDD-assigned names in the SEDD DED. Data requesters are urged to carefully check the SEDD DED to ensure that the needed data element is not already present in the dictionary under another name.

  NOTE: Please check with the SEDD Program Manager for any assistance regarding the addition of new data elements.

- Names should be spelled-out and each word should have its first letter capitalized (e.g., 'ResultUnits') using the Upper Camel Case (UCC) convention. Excessive abbreviation should be avoided.

Data Element Values

- A data element value can contain any string of characters allowed in a line, as restricted by rules defined by the data requester. A value is the data delivered by a data element.

  NOTE: Part of the definition of a name is a possible restriction on the format of values associated with it, such as numeric or date formats. Section 2.3 defines these formats.

- By SEDD rules, the length of the value is specified by the data requester.

- The value begins immediately after the starting data element tag name. In particular, any spaces after this starting tag and before the ending tag are part of the value.

- A value can be null (contain no characters). Null values would be reported on a single line using the following format: optional leading spaces; a less than sign, followed by the data element tag name, followed by a forward slash, followed by the greater than sign (e.g., <DataElementTag/>).

Nodes

- A node element consists of an opening node tag and a closing node tag which are located on separate lines within the EDD. The opening node tag is a less than sign followed by a node tag name followed by a greater than sign (e.g., <NodeTag>). A closing node tag name must also be used later in the EDD and consists of a less than sign, followed by a forward slash, followed by the same node tag name, followed by the greater than sign (e.g., </NodeTag>).

- The first node encountered in the EDD is referred to as the root node. This node can only be used once. All data delivered in the EDD must fall within the opening root node (e.g., <Header>) and the closing root node (e.g., </Header>).

Node Element Names

- A node element name is a string of alphanumeric (A-Z, a-z, 0-9) characters.

- A name should be limited to a maximum of 30 characters.

- Character case is significant in names. The first letter of all words used in node element names must be capitalized. Example: SamplePlusMethod would be acceptable while SampleplusMethod would not.

- The following are SEDD Specification-defined node element names and a description of the intended use of the associated elements.

Analysis	Describes one complete sequence of events, from taking a sample aliquot and preparation through measurement, defined as part of one method.
AnalysisGroup	Links several analyses used to compute results for one method.
Analyte	Per-analyte data from one analysis or one group of analyses.
AnalyteComparison	Describes the effects of potentially interfering analytes on a peak (e.g., ICP Interelement Correction factors).
AnalyteGroup	Links several analytes used to compute results for another analyte.
Handling	Describes any manipulation of the sample prior to taking an aliquot (e.g., filtering, ashing, leaching).
Header	Carries information for the reader/parser of an EDD data stream.
InstrumentQC	Describes QC analyses that do not yield results (concentrations). These data are usually about instrument or process QC (e.g., initial and continuing calibration data).
Peak	Reports an actual measurement made during one analysis of one analyte.
PeakComparison	Describes cross-peak comparisons (e.g., abundance ratios and inter-peak resolutions).
PeakReplicate	Reports data when multiple peak measurements are made (e.g., multiple exposure readings).
PreparationPlusCleanup	Describes a preparation or cleanup as part of an analysis or an independent method.
ReportedResult	Contains the actual results of a method (analyte identification and computed final results).
SamplePlusMethod	Contains data about the characteristics of one sample analyzed under the criteria of one method.

- Implementation-defined nodes are not allowed.

- Node elements should be spelled-out and each word should have its first letter capitalized (e.g., 'ReportedResult') using the Upper Camel Case (UCC) convention. Excessive abbreviation should be avoided.

Node Elements

- A node definition begins at a node definition line and continues until the closing node definition line is encountered.

- The contents of a node are the node definition line that starts it, all data elements and other nodes within it, and the node definition line that ends it. In particular, blank and comment lines (<!-- Comments -->, as opposed to Comment data elements) are not part of any node.

- Most data element names can only be used in specific nodes (see Appendix A). However, some data element names may appear in more than one node, possibly with slightly different definitions.

- A data element name may not appear more than once in any given node.

- With the exception of data elements denoted as required by the SEDD specification (see Section 2.2) and those specified as required by a SEDD-compatible EDD implementation, no data element is required to appear in an EDD.

- SEDD specification has no rules restricting the order of data elements in nodes and data requesters cannot add any such rules.

2.1.3 SEDD Hierarchy, Header and Dependent Data Element Syntax

- SEDD specification nodes are arranged in a hierarchy. Figure 7 defines this hierarchy. As implied by the term hierarchy, nodes at any level can repeat as many times as needed.

- Subject to the rules associated with the hierarchy, there are no restrictions on the number, type, and order of nodes in an EDD and a data requester cannot add any. For example, an empty EDD is valid, if not useful. Data requesters can choose to ignore nodes they do not recognize or which are of no interest to them.

- Parent nodes with identical content cannot be repeated. For example, a SamplePlusMethod node could not be repeated before each individual ReportedResult node associated with it. It is more efficient to only have one SamplePlusMethod node for all associated ReportedResults' nodes.

  NOTE: Data requesters need to check an EDD for global consistency of all potentially redundant data.

Header Nodes

- A Header node is always the first node in an EDD. It provides the information (e.g., EDD version and implementation identification) needed to identify and process the EDD reliably.

- Each Header node can only refer to and report data based on a single DTD or Schema. Data for multiple methods can be reported under a single Header node provided that the same DTD or Schema can be used by each method.

- An EDD (e.g., a disk file), may have multiple Header nodes as long as each header is unique. Data processing must start fresh at each new Header node.

- If two valid EDDs are concatenated (i.e., two data files with different Header nodes are appended in one file), the result is one valid EDD.

Other Nodes

- In an EDD, all other nodes must be preceded by at least one node at each higher level. Figure 7 defines the dependencies between these nodes. For example, the SamplePlusMethod node is dependent on a Header node.

- Any given node is said to be associated with the closest preceding node at each higher level in an EDD.

- The pattern of associated higher-level nodes must match the same pattern as in Figure 7. For example, in order to have a valid ReportedResult node in an EDD, it must be preceded by a SamplePlusMethod node and a Header node, as shown in Figure 7. There cannot be an InstrumentQC node between the ReportedResult and SamplePlusMethod nodes. Otherwise, the ReportedResult node would not relate to a sample. This concept is called nesting. All nodes must be properly nested.

- When more than one node of a higher level and identical type precedes another node, the dependent node is related to the nearest higher order node as shown in the following example:

<Header>
<!-- *Header Node - Beginning Tag* -->
    <SamplePlusMethod>
    <!-- *First SamplePlusMethod node - Beginning Tag* -->
        <ReportedResult>
        </ReportedResult>
        <!-- *Associated with the first SamplePlusMethod node* -->
        <ReportedResult>
        </ReportedResult>
        <!-- *Associated with the first SamplePlusMethod node*-->
    </SamplePlusMethod>

```
        <!-- First SamplePlusMethod node - Ending Tag -->
        <SamplePlusMethod>
        <!-- Second SamplePlusMethod node - Beginning Tag  -->
            <ReportedResult>
            </ReportedResult>
            <!-- associated with the second SamplePlusMethod node -->
        </SamplePlusMethod>
        <!-- Second SamplePlusMethod node - Ending Tag -->
</Header>
<!-- Header node - Ending Tag -->
```

- While the order of nodes at different levels is important in the EDD, the order of nodes at the same level is not. This is shown in the following two examples, where the ReportedResult and Analysis Nodes are at the same level:

EXAMPLE 1:

```
<Header>
        <SamplePlusMethod>
            <ReportedResult>
            </ReportedResult>
            <ReportedResult>
            </ReportedResult>
            <ReportedResult>
            </ReportedResult>
            <Analysis>
            </Analysis>
        </SamplePlusMethod>
</Header>
```

EXAMPLE 2:

```
<Header>
        <SamplePlusMethod>
            <ReportedResult>
            </ReportedResult>
            <Analysis>
            </Analysis>
            <ReportedResult>
            </ReportedResult>
            <ReportedResult>
            </ReportedResult>
        </SamplePlusMethod>
</Header>
```

These two examples are identical in meaning, but differ in order. All of the ReportedResult nodes in both examples associate, hierarchically, to the preceding SamplePlusMethod node. The Analysis node in each example also belongs, hierarchically, to the same SamplePlusMethod node. The Analysis node does not modify any of the ReportedResult nodes in the second example, even though it precedes it. This is because the Analysis node is not of a higher level than the ReportedResult node.

2.1.4 XML File Syntax

- Data will be delivered from data generators to their data requesters as XML documents. These XML documents will be both well-formed and valid data files.

- A well-formed XML document follows all of the official XML rules as defined by the World Wide Web Consortium (W3C). Many of these rules have already been discussed in previous sections. Data requesters are encouraged to acquaint themselves further with XML using some of the many resources available at this time.

- A valid XML document is a document that follows the Document Type Definition (DTD) file or XML Schema as established by the data requester. The DTD or XML Schema gives the hierarchical structure and states what data elements will be applicable for a given application.

2.2 TYPES OF REQUIRED DATA ELEMENTS

A few data elements are required to have non-null values to properly and uniquely identify the reported data. If a node is not required, any data element it may contain is not required. The SEDD Specification has four types of required data elements:

- Required Data Elements (see Section 2.2.1)
- Conditionally Required Data Elements (see Section 2.2.2)
- Data Elements Required for Traceability (see Section 2.2.3)
- Data Elements Required for Portability (see Section 2.2.4)

Figure 9 identifies the data elements required in a Header node. Figure 10 and Figure 11 identify the data elements required for SamplePlusMethod and InstrumentQC nodes, and related nodes, respectively. Footnotes in these figures classify each data element using the above types. These figures are presented as a useful summary of the SEDD Specification hierarchy and key data elements. Appendix A contains the current specifications for data elements required for SEDD.

FIGURE 9. SEDD SPECIFICATION REQUIRED DATA ELEMENTS: Header NODE

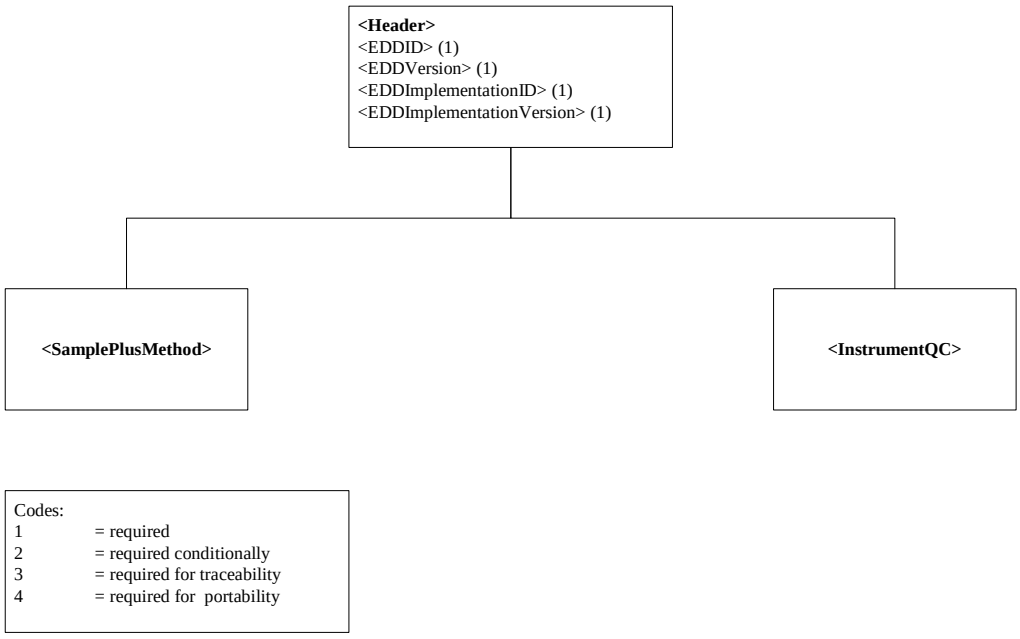


FIGURE 10. SEDD SPECIFICATION REQUIRED DATA ELEMENTS: SamplePlusMethod AND RELATED NODES

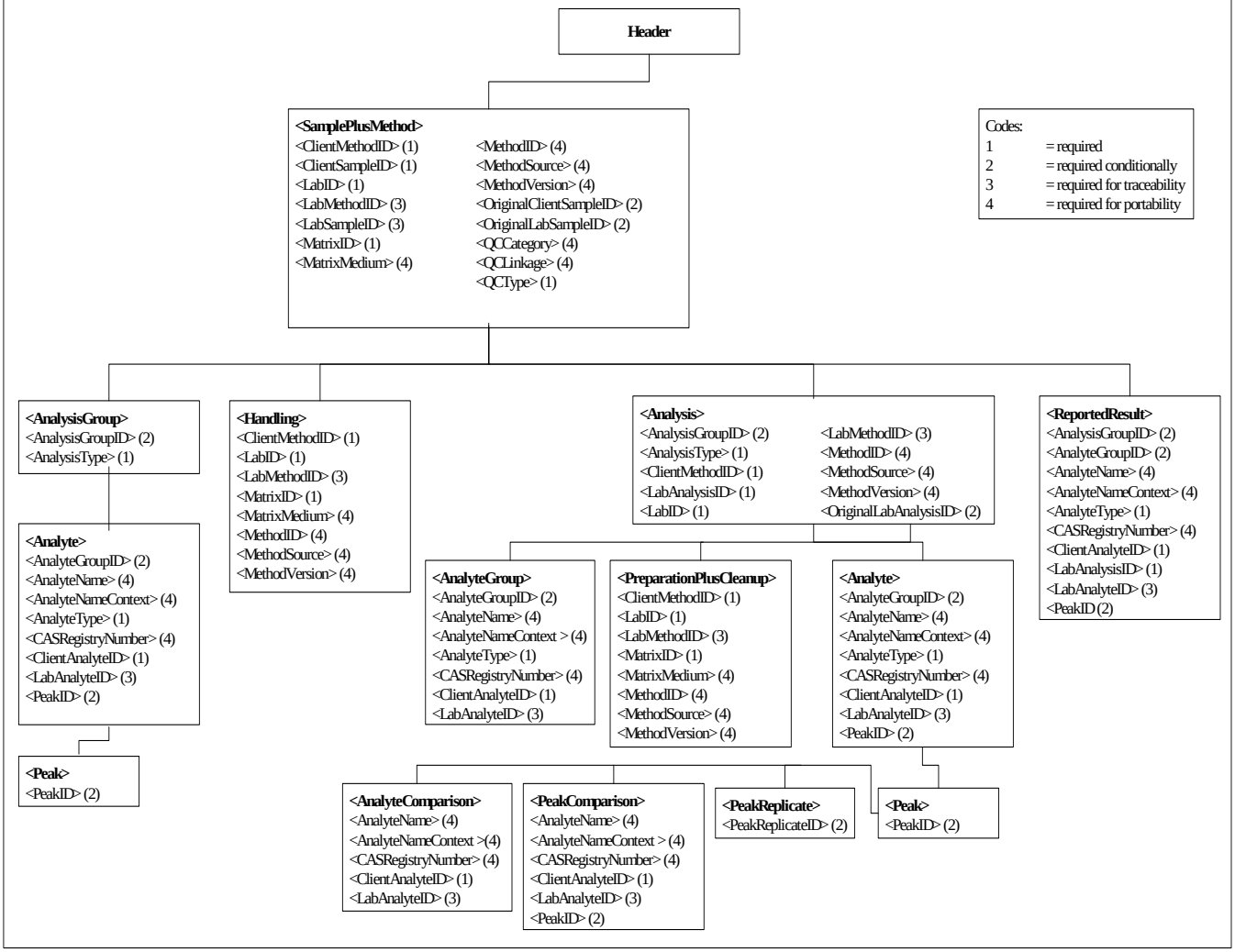
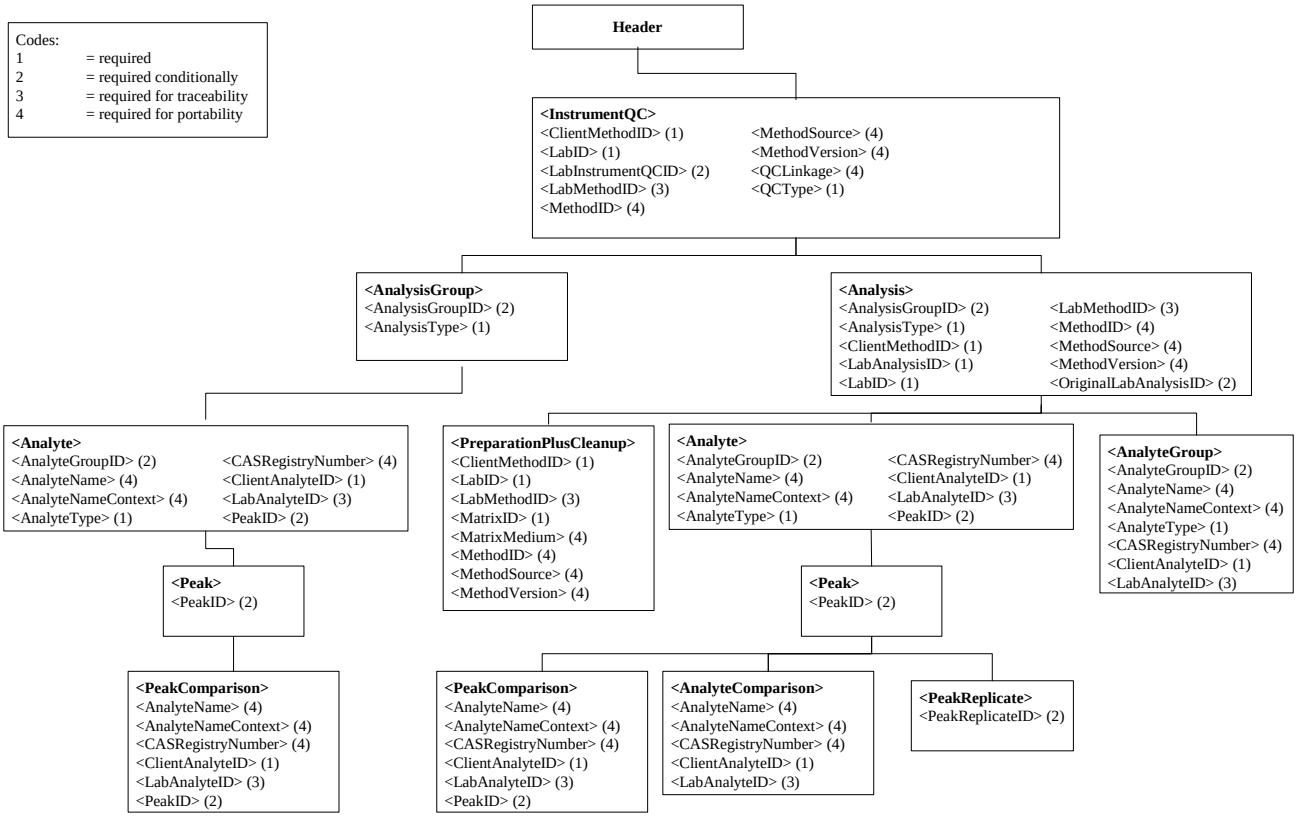


FIGURE 11. SEDD SPECIFICATION REQUIRED DATA ELEMENTS: InstrumentQC AND RELATED NODES



2.2.1 Required Data Elements

Required data elements are the basis for reliable identification of EDD data by the data requestor. Examples include QCType, AnalysisType, and AnalyteType. These data elements MUST be present in a valid EDD.

2.2.2 Conditionally Required Data Elements

Conditionally required data elements also are the basis for identification and linkage of data, but are not required in all cases. Examples include OriginalClientSampleID, required for certain QC samples; LabInstrumentQCID, required when reporting Instrument QC samples; and AnalysisGroupID, required when multiple analyses are used to generate a single result. Based on rules specified for each data element in Appendix A, these data elements must be present in a valid EDD.

2.2.3 Data Elements Required for Traceability (Including QC)

The SEDD Specification has standard data elements to use when traceability back into the laboratory's process is desired. Examples include LabSampleID and LabAnalyteID. These data elements are not formally required by the SEDD Specification, but are suggested as likely to be required in the instructions from the data requester to the data generator.

2.2.4 Data Elements Required for Portability (Including QC)

The SEDD Specification has standard data elements to use when portability of data across sites and/or data requesters is desired. Examples include AnalyteName, QCCategory, and QCLinkage. These data elements are not required by the SEDD Specification, but are suggested as likely to be required in the instructions from the data requester to the data generator.

NOTE 1: The SEDD Specification requires the use of all required and conditionally required data elements. Data requesters may require additional data elements based on program needs.

NOTE 2: For consistency, the SEDD Specification has required fields in many nodes that serve to uniquely identify the data. For example, the PeakID data element in a Peak node is a conditionally required data element. It is needed to uniquely identify the peak in a GC/MS analysis. A reasonable value for the PeakID data element is the mass of the peak quantitated.

NOTE 3: There are some situations where no unique values exist for the required data element. For example, the ClientAnalyteID data element in a ReportedResult node is a required data element. However, there would be no real ClientAnalyteID for a Tentatively Identified Compound (TIC) as might be reported during a GC/MS analysis. In this situation the data requester must define how this data element will be used for TICs (i.e., ClientAnalyteID = Unknown_Hydrocarbon01).

As noted in the beginning of this section, the data requester must specify the values for these required data elements if SEDD required valid values are not available (see Appendix B).

2.3 DATA ELEMENT VALUE FORMATS

There are five different types of data that can be entered into a given data element. These five types are the data element value formats. These formats are Text, Identifier, Limited List, Numeric, and Date. The Data Element Dictionary (Appendix A) gives the format for each data element.

The examples given below are a partial list of the most commonly used formats. The data requester must specify the exact format of each data element depending on program needs.

2.3.1 Text

- Text format allows any value consistent with the syntax of a line. Case is significant in Text format so that "Test", "TEST", and "tEst" are all different values for a Text data element. If data is to be parsed to a database, this may be less of an issue.

- Leading and trailing spaces are significant in Text data element values.

2.3.2 Identifier

- Identifier format is a restricted version of Text format. Typically, these values use a restricted character set. The data requester should provide instructions for each data element with this format.

- A suggestion is that characters be restricted to alphabetic characters, digits, the underscore character, and the dash character and that case not be treated as significant. Identifiers can have further restrictions, such as check digits in CAS Numbers or certain required patterns of alphabetic and numeric characters.

2.3.3 Limited List

- Limited list format is another restricted version of Text format. Typically, these values are limited to a short list of valid values. The DTD or Schema should specify valid values for each data element with this format. Valid values may depend on the value of other data elements. For example, valid values for the ResultUnits data element would typically depend on the MatrixID data element. Appendix B of the SEDD Specification contains a list of valid values for the data elements that use this format.

2.3.4 Numeric

- Numeric format includes any of the following formats:

  - **Integer**: leading spaces, a minus sign, a string of digits and trailing spaces, with all subparts optional. The value is interpreted as a whole number.

  - **Decimal**: leading spaces, a minus sign, a string of digits, a decimal point, another string of digits and trailing spaces, with all parts but the decimal point optional. The value is interpreted as a real number.

  - **Exponential**: leading spaces, a minus sign, a string of digits, a decimal point, another string of digits, a string of spaces, the letter "E" or "e", a string of spaces, a plus or minus sign, a string of digits and trailing spaces.

- A data requester should be prepared to accept any of the possible numeric formats. A data requester must be prepared to read all three formats within the same data stream. However the data requester can specify rules for this format.

NOTE 1: The same number can be expressed in an EDD in the following ways and all are allowable:

12345	Integer format
12345.000	Decimal format
12345E 0	Exponential format with a zero exponent

If any of the above values are preceded by a minus sign, they are valid numbers.

NOTE 2: Legal values for zero include, but are not limited to:

0
0.0
0.0e0

- The SEDD Specification by itself does not limit the precision of numbers. Data requesters should specify minimum acceptable significant digits and/or precision for every numeric data element used. It is permissible for data generators to report greater precision than specified in a DTD or Schema.

- The SEDD Specification does not specify rounding rules. This has to be specified by the data requester.

- A null value for a numeric data element represents a totally blank value, which is not considered equivalent to the number zero.

2.3.5 Date

- The date format includes, but is not limited to, the following:

MM/DD/YY HH:mm:SS
MM/DD/YY HHmmSS
MM/DD/YYYY HH:mm:SS
MM/DD/YYYY HHmmSS
MMDDYY HH:mm:SS
MMDDYY HHmmSS
MMDDYYYY HH:mm:SS
MMDDYYYY HH:mm:SS
YYMMDD HH:mm:SS
YYMMDD HHmmSS
YYYYMMDD HH:mm:SS
YYYYMMDD HHmmSS

Where:

MM = month
DD = day
YY = last two digits of the year
YYYY = four digits of the year
HH = hour in 24 hour format
mm = minutes
SS = seconds

- The time portion of a date, with any preceding space, is optional unless required by a data requester for specific data elements. The seconds portion of the time, with any colon, is similarly optional.

- All date format values in the data stream following a Header parent data element must be in the same date format. This format is specified by the DateFormat data element in the Header node (see Appendix A).

  NOTE: If no date format is specified, the default format YYYYMMDD HHmmSS is used.

3.0 CONCEPTS AND RELATIONSHIPS

Methods used for the analysis of environmental samples are complicated procedures that include sample preparation (digestion or extraction), cleanup of sample extracts prior to analysis, and the actual instrumental analysis. There are many associations for a single sample analysis that must be captured in a unique way if the whole process has to be reconstructed. This is especially important if the data requester wants to clearly link all Quality Control (QC) data with the sample result.

Describing associations among data elements is essential for an EDD to be useful. Because of the variety of batches, groupings, linkages, and comparisons needed for each data requester, specifying a DTD or Schema and the resultant EDD that correctly describes a method can be difficult. Section 3.1 describes some of the concepts underlying the SEDD Specification, as follows:

- Samples (see Section 3.1.1)
- Instrument QC (see Section 3.1.2)
- Method (see Section 3.1.3)
- Method QC Sample (see Section 3.1.4)
- Analysis (see Section 3.1.5)
- Results (see Section 3.1.6)

Section 3.2 provides a discussion of the more complicated analytical relationships within the SEDD Specification. Batches typically involve linking the Method Quality Control samples (e.g., Laboratory Control Samples, Method Blanks, Matrix Spikes) analyzed to the regular field samples. Analysis Groups (under the SamplePlusMethod node) are used whenever the final reported result for a field sample is computed from more than one separate analysis result (reported as intermediate results of the sample aliquot) as in the case of Method of Standard Additions for Inorganic Analysis. Analysis Groups (under the InstrumentQC node) are used whenever multi-point initial calibration curves are determined. QC Linkages specify which batch types are used to link QC samples to regular field samples. Comparisons are used to relate individual measurements to other reference measurements for Quality Control purposes. The remainder of Section 3 is divided into the following sections:

- Batches (see Section 3.2.1)
- Analysis Groups (see Section 3.2.2)
- Analyte Groups (see Section 3.2.3)
- QC Categories and QC Linkages (see Section 3.2.4)
- Comparisons - Peak and Analyte (see Section 3.2.5)

This section makes liberal use of examples drawn from actual EDDs developed under this specification. Examples in this section have been formatted to aid readability. Formatting conventions need not be adopted when preparing an EDD. While reviewing the following concepts and analytical relationships, keep Figure 7 handy as a reference for the hierarchy of the nodes in the SEDD Specification.

3.1 CONCEPTS

This section provides a discussion of the fundamental concepts within the SEDD Specification. This includes a discussion of the various types of samples likely to be encountered, the methods used for their analysis, and the reporting of results.

3.1.1 Samples

A sample is defined in the SEDD Specification as any substance given a sample number or identifier. This includes ones sent by a data requester to the data generator (e.g., an analytical laboratory) for analysis and certain Method QC samples [e.g., Laboratory Control Samples (LCSs), Method Blanks] generated by the data generator (see Section 3.1.4). In a SEDD Specification-compatible EDD, a sample is reported in SamplePlusMethod and related nodes (i.e., all nodes located below the SamplePlusMethod node in the hierarchy) based on the application of one primary analysis method (see Figure 7).

NOTE: The SEDD Specification does not have a per-sample node. This is because the notion of what a "sample" is as reported in an EDD is ambiguous. For example, the sample might be what is identified by the following data elements within the SamplePlusMethod node: ClientSampleID, FieldSampleID, or LabSampleID. In real cases involving multiple methods, handlings, and reanalyses, these data elements may have different values.

The SEDD Specification uses the SamplePlusMethod node as its highest-level of reporting for samples. The association between the method and a sample must be clear to both the data requester and the data generator. This association is NOT necessarily provided as part of the DTD or Schema and must be provided by the data requester in their instructions to the data generator.

Example Sample Definition:

Some programs may report a dilution of a sample separately from its original (i.e., in two separate SamplePlusMethod nodes) with a distinct ClientSampleID data element and a separate set of results in each SamplePlusMethod node. When a program reports a dilution of a sample separately from its original in a SEDD Specification-compatible implementation, then two separate SamplePlusMethod nodes will be used, one node to report the results of the original sample and the second node to report the results of the diluted sample. There will be only one Analysis node under each SamplePlusMethod node. In contrast, other programs may consider dilutions to be internal to the methods and only one set of selected results per original sample be reported from the multiple analyses (i.e., in only one SamplePlusMethod node). When a program reports a dilution of a sample with only a single set of selected results per original sample, then separate Analysis nodes under one SamplePlusMethod node will be used.

The following information must also be defined in the data requester instructions:

- Samples should have a well-defined matrix identified by a MatrixID data element.

- In each SamplePlusMethod node the primary analytical method applied is identified by a MethodID (or ClientMethodID or LabMethodID) data element.

3.1.2 Instrument QC

- Any analysis not reportable as part of the data for a sample (as defined in Section 3.1.1) is generically called instrument QC and is reported in InstrumentQC and related nodes (all nodes located below the InstrumentQC node in the hierarchy - see Figure 7).

- One InstrumentQC node conveys information associated with one QC measurement, classified by a QCType data element (e.g., calibrations and instrument blanks). Analyses not directly associated with the primary analytical instrument (e.g., GPC calibrations) are also included.

- Typically, only one analysis is reported under one InstrumentQC node. However, some types of instrument QC require multiple analyses to assess a particular performance characteristic. For example, many initial calibrations require multiple analyses to assess linearity of instrument response. In this case, multiple analyses are reported under one InstrumentQC node, using AnalysisGroup nodes to report values computed from the multiple analyses. Each InstrumentQC node must be uniquely identified through the use of the LabInstrumentQCID data element.

Listed below is an example of how an XML file would look regarding the reporting of initial calibration data which has multiple analyses. This example, derived from the analysis of semivolatile organics using SW-846 Method 8270, illustrates how analyses in one initial calibration might be reported using ONLY the Analysis and AnalysisGroup nodes (does not go below into the Analyte nodes). This example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- One instrument QC event: -->

<InstrumentQC>
        <!-- identification -->
        <ClientMethodID>8270C</ClientMethodID>
        <LabInstrumentQCID>ICAL-1</LabInstrumentQCID>
        <!-- classification -->
        <QCType>Initial_Calibration</QCType>

        <!-- 1st Analysis Node and data elements -1st standard used for initial calibration. -->

        <Analysis>
            <!-- identification -->
            <LabAnalysisID>SV202</LabAnalysisID>
            <AnalysisGroupID>ICAL01</AnalysisGroupID>
            <!-- classification -->
            <AnalysisType>RRF-010</AnalysisType>
            <!-- typical analysis data -->
            <AnalyzedDate>11/20/1997 12:50</AnalyzedDate>
            <InstrumentID>GCMS1</InstrumentID>
        </Analysis>

        <!-- 2nd Analysis Node and data elements, 2nd standard used for init cal. - ->

        <Analysis>
            <LabAnalysisID>SV203</LabAnalysisID>
            <AnalysisGroupID>ICAL01</AnalysisGroupID>
            <AnalysisType>RRF-020</AnalysisType>
            <AnalyzedDate>11/20/1997 13:50</AnalyzedDate>
            <InstrumentID>GCMS1</InstrumentID>
        </Analysis>

        <!-- 3rd Analysis Node and data elements -3rd standard used for init cal.  -->

        <Analysis>
            <LabAnalysisID>SV204</LabAnalysisID>
            <AnalysisGroupID>ICAL01</AnalysisGroupID>
            <AnalysisType>RRF-050</AnalysisType>
            <AnalyzedDate>11/20/1997 14:50</AnalyzedDate>
            <InstrumentID>GCMS1</InstrumentID>
        </Analysis>

        <!-- 4th Analysis Node and data elements -4th standard used for initial calibration. -->

        <Analysis>
            <LabAnalysisID>SV205</LabAnalysisID>
```

```
        <AnalysisGroupID>ICAL01</AnalysisGroupID>
        <AnalysisType>RRF-080</AnalysisType>
        <AnalyzedDate>11/20/1997 15:50</AnalyzedDate>
        <InstrumentID>GCMS1</InstrumentID>
    </Analysis>

    <!-- 5th Analysis Node and data elements -5th standard used for initial calibration. -->

    <Analysis>
        <LabAnalysisID>SV206</LabAnalysisID>
        <AnalysisGroupID>ICAL01</AnalysisGroupID>
        <AnalysisType>RRF-160</AnalysisType>
        <AnalyzedDate>11/20/1997 16:50</AnalyzedDate>
        <InstrumentID>GCMS1</InstrumentID>
    </Analysis>

<!-- This concludes the Analysis Nodes. -->

    <AnalysisGroup>
        <AnalysisGroupID>ICAL01</AnalysisGroupID>
<!-- Mean RRFs and RSDs would go in data elements in the Analyte nodes nested here: -->
    </AnalysisGroup>

</InstrumentQC>
```

3.1.3 Method

- A method corresponds to a defined process for the identification and quantitation of selected analytes. The analyte list for a method often corresponds to compounds or substances measurable after one analysis on one analytical instrument.

- A method should include specifications for the type, frequency, and performance criteria for QC samples.

- Details of a method can be client-specific. The following SEDD Specification data elements, ClientID, ClientMethodID (for the instrumental analysis), and MatrixID, are used to identify, not describe, a method. This allows the reader of an EDD to look up in their own database whatever method characteristics are needed to correctly process the data. Some programs, such as the Contract Laboratory Program (CLP), define the method to include all of the sample processing steps. Other programs, such as SW-846, will define unique methods for each of these steps.

- The SEDD Specification defines four types of activities associated with applying a method to a sample: characterization; handling; preparation; and instrumental analysis.

- **Characterization**

Characterization applies to the sample as received by the data generator. This includes recording or measuring characteristics such as weight, color, texture, temperature, moisture, and pH. All such data is reported in the SamplePlusMethod node using various data elements (e.g., pH, Clarity and Turbidity). Characterization might include a screening process to determine which variant of a method to apply (e.g., determining the level for Organics).

Here are two examples for reporting Characteristics.

Example 1 - Use of a Screening Method

In a complex case (e.g., a GC screen prior to a GC/MS analysis) the details of the screen could be reported under a separate SamplePlusMethod node with a different ClientMethodID. However, it is not common to report this level of detail.

Example 2 - Reporting of Percent Moisture (% Moisture)

In some cases, a characteristic such as Percent Moisture can be reported in one of two possible ways. The first, and most common approach is to treat Percent Moisture as a characteristic property of the sample itself and report it using the PercentMoisture data element in the same SamplePlusMethod node as the sample. The second approach is to treat Percent Moisture as a separate test entirely. For this situation this test would then be reported in a separate SamplePlusMethod node with a separate Results node for Percent Moisture.

- **Handling**

Handling applies to any manipulation of the sample prior to taking an aliquot for preparation/analysis. Examples include filtering, decanting, drying, grinding, ashing, and leaching (TCLP in particular). It is common to further characterize the sample (e.g., by weight) after each handling. All such data is reported in a Handling node under the appropriate SamplePlusMethod node. Each Handling node is characterized by a ClientMethodID or HandlingType data element.

Many methods have no handling stage and no Handling node is required in these cases. Less commonly, more than one handling is done, so more than one Handling node is required.

Example:

A sample is dried and the weight is recorded in one Handling node using one HandlingType data element. In a subsequent step, the dried sample is ashed and the weight after ashing is recorded in a second Handling node with a different HandlingType data element.

- **Preparation**

  Preparation applies to all processing done to an aliquot prior to instrumental analysis. The details of preparation might involve many steps (e.g., taking an aliquot, extraction, and cleanup). Most methods have a primary processing step, such as chemical extraction or separation that are part of the analysis method. The preparative steps could also be described in a separate method. The specific preparation details would normally be captured in a separate PreparationPlusCleanup node, not in the Analysis node.

  A few values, such as aliquot size, dilution, and yield (radiochemistry), are often important to report for each preparation. These data can go in an Analysis or a PreparationPlusCleanup node. For analyses that require a minimum or no preparative steps, no PreparationPlusCleanup nodes would be needed.

  PreparationPlusCleanup nodes are used to report specific preparation steps, especially when a separate method is used to describe this activity. These are similar to Handling nodes in that there might be none, one, or several, depending on the method. The difference is that a PreparationPlusCleanup node applies to one aliquot used in a preparation, while a Handling node applies to one sample prior to taking an aliquot.

  The ClientMethodID or CleanupType data element in the PreparationPlusCleanup node is used to characterize the cleanup.

- **Instrumental Analysis**

  Instrumental analysis, more generally called the determinative step, is where measurements are made for a list of analytes. Values such as instrument identification and date analyzed are reported in the Analysis node. Analyte-specific values from this analysis are reported in Analyte nodes under the Analysis node.

  If the analytical technique involves measurements of multiple peaks per analyte (e.g., GC/MS mass spectra, multi-component GC analytes, ICP emission spectra), details would normally be reported in Peak nodes under each Analyte.

  The following example, derived from the SW-846 Semivolatile data for Method 8270, illustrates how the various activities in one method are reported in XML format. This example uses the following Nodes: SamplePlusMethod; Handling; Analysis; and PreparationPlusCleanup. The Analyte node is not used. This example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- identify and characterize one sample with one method applied -->

<SamplePlusMethod>
        <!-- identification -->
        <ClientMethodID>8270C</ClientMethodID>
        <ClientSampleID>S001</ClientSampleID>
        <CollectedDate>12/03/1997</CollectedDate>
        <MatrixID>Soil</MatrixID>
        <LabReceiptDate>12/04/1997</LabReceiptDate>
        <LabSampleID>M971204-004</LabSampleID>
        <!-- a typical characterization -->
        <Temperature>4.4</Temperature>
        <TemperatureUnits>C</TemperatureUnits>
        <!-- the 'as received value' -->
        <PercentSolids>65</PercentSolids>

        <!-- a typical handling -->

        <Handling>
            <!-- classification -->
            <HandlingType>Decanted</HandlingType>
            <!-- the 'after decanting' value -->
            <PercentSolids>82</PercentSolids>
        </Handling>

        <!-- instrumental analysis -->

        <Analysis>
            <!-- identification -->
            <LabAnalysisID>SV422</LabAnalysisID>
            <!-- analytical method used -->
            <ClientMethodID>8270C</ClientMethodID>
            <!-- typical analysis values -->
            <AnalyzedDate>12/24/1997 14:38</AnalyzedDate>
            <DilutionFactor>1</DilutionFactor>
            <!-- the results will be reported on a 'dry-weight' basis -->
            <ResultBasis>Dry</ResultBasis>

            <!-- a typical preparation -->

            <PreparationPlusCleanup>
                <!-- preparation method used -->
                <ClientMethodID>3550B</ClientMethodID>
                <!-- typical preparation values -->
                <AliquotAmount>30.0</AliquotAmount>
                <AliquotAmountUnits>g</AliquotAmountUnits>
```

```
            <FinalAmount>1.00</FinalAmount>
            <FinalAmountUnits>mL</FinalAmountUnits>
            <PreparedDate>12/07/1997</PreparedDate>
        </PreparationPlusCleanup>

    <!-- a typical cleanup event -->

        <PreparationPlusCleanup>
            <!-- cleanup method used -->
            <ClientMethodID>3640A</ClientMethodID>
            <!-- typical cleanup values -->
            <CleanedUpDate>12/07/1997</CleanedUpDate>
        </PreparationPlusCleanup>
    </Analysis>
</SamplePlusMethod>
```

3.1.4 Method QC Sample

- A QC sample suitable for reporting in SamplePlusMethod nodes is called a method QC sample (e.g., Laboratory Control Sample, Method Blank, Matrix Spike). Its matrix should be well defined. Results are normally reported in the same units (exceptions would include serial dilutions and post-digestion spikes) for associated regular samples. Method QC samples are a component of performance-based evaluation of analytical data.

- No handling may be associated with Method QC samples. However, they are prepared in the same manner as regular samples.

The following example, derived from the SW-846 Method 8270 semivolatile data, shows how results and QC values for a spiked analyte are reported for a matrix spike sample using the SamplePlusMethod and ReportedResult nodes. This example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- One (method QC) sample and method: -->

<SamplePlusMethod>
        <!-- identification -->
        <ClientMethodID>8270C</ClientMethodID>
        <MatrixID>Water</MatrixID>
        <LabSampleID>M971204-005</LabSampleID>
        <OriginalClientSampleID>S001</OriginalClientSampleID>
        <!-- classification -->
        <QCType>Matrix_Spike</QCType>
        <QCCategory>Spike</QCCategory>
        <QCLinkage>PreparationBatch</QCLinkage>

        <!-- Typical spiked analyte values: -->

        <ReportedResult>
            <ClientAnalyteID>108-95-2</ClientAnalyteID>
            <AnalyteName>Phenol</AnalyteName>
            <AnalyteType>Spike</AnalyteType>
            <ExpectedResult>75</ExpectedResult>
            <ExpectedResultUnits>ug/L</ExpectedResultUnits>
            <Result>67.3</Result>
            <ResultUnits>ug/L</ResultUnits>
            <PercentRecovery>71.1</PercentRecovery>
       </ReportedResult>
</SamplePlusMethod>
```

3.1.5 Analysis

The SEDD Specification defines an analysis as one complete sequence of events starting with an aliquot or prepared sample, perhaps involving preparation, and including an instrumental analysis. This information would be captured in the Analysis node and related nodes (all nodes located below the Analyses node in the heirarchy - see Figure 7). An analysis may be part of a method applied to a sample or part of an instrumental QC process. Thus Analysis nodes are present under both the SamplePlusMethod node and the InstrumentQC node.

3.1.6 Results

- The final results of a method are always reported in ReportedResult nodes. Intermediate results underlying the final results are reported in Analyte nodes, however, final results can also be reported in Analyte nodes. These final results take into consideration any dilutions or Percent Moisture calculations that would be needed.

  NOTE: In the simple case of only one analysis per method, the same values could be reported in both ReportedResult and Analyte nodes. The ReportedResult nodes are required. The Analyte nodes are optional and might be used to report intermediate

and final measurements or final measurements from both a primary and confirmation analysis.

- The SEDD Specification distinguishes between the result of a method (e.g., reported concentration of benzene of 250 ug/L in a groundwater sample by Method 8260) and the result of an analysis (e.g., serial dilution for metals). Method results are reported in ReportedResult nodes directly under the SamplePlusMethod node. Analysis results can be reported in Analyte nodes under Analysis nodes or in Peak nodes under Analyte nodes.

- Because InstrumentQC does not have a method-like result, no ReportedResult nodes are used.

- Certain analytes in certain methods are always measured on a per-analysis, not per-method, basis (e.g., surrogates and internal standards). They should be reported in Analyte nodes, not under ReportedResult nodes. For example: in the analysis of pesticides there are separate surrogate recoveries computed for each column and surrogate analyte, which must be reported in Analyte nodes.

- Whenever more than one analysis is performed, either the data element LabAnalysisID or the data element AnalysisGroupID must be used. The data element LabAnalysisID must be used to link each reported result to the SINGLE underlying Analysis data used to produce that result. The LabAnalysisID data element must be present in both the Analysis node and ReportedResult node. If each reported result is calculated from MULTIPLE underlying Analysis data, then the AnalysisGroupID data element must be used. See Section 3.2.2 for a complete discussion of analysis groups.

- There are three ways to report sample reanalysis data:

  1) If the laboratory is to report two sets of results, two SamplePlusMethod nodes should be used.

  2) If the laboratory is to pick one “best” result for each analyte, use one SamplePlusMethod node. The LabAnalysisID data element would be used to link each result to the correct underlying analysis.

  3) If the data requester requires all potential results in addition to the selected result(s), use Analyte nodes to report them under each Analysis node.

     The following example, derived from the SW-846 Inorganics data, illustrates how both final and intermediate results can be reported for one soil sample. This example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- Regular Soil Sample - ICP Method -->

<SamplePlusMethod>
        <ClientMethodID>6010C</ClientMethodID>
        <ClientSampleID>BH-0723Q</ClientSampleID>
        <LabSampleID>980917-006</LabSampleID>
        <MatrixID>Soil</MatrixID>

        <Analysis>
            <LabAnalysisID>ICP-CD1-981103-49</LabAnalysisID>
            <ClientMethodID>6010C</ClientMethodID>

            <Analyte>
                <ClientAnalyteID>7440-43-9</ClientAnalyteID>
                <!-- Intermediate result in water units.  Required by some -->
                <!-- methods if a post digestion spike and/or serial dilution is -->
                <!-- also done. -->
                <IntermediateResult>7.2</IntermediateResult>
                <IntermediateResultUnits>ug/L</IntermediateResultUnits>
            </Analyte>
        </Analysis>

        <!-- Final Results -->

        <ReportedResult>
            <ClientAnalyteID>7440-43-9</ClientAnalyteID>
            <LabAnalysisID>ICP-CD1-981103-49</LabAnalysisID>
            <LabQualifiers></LabQualifiers>
            <!--This is the final reported result in soil units for the soil sample.-->
            <Result>0.76</Result>
            <ResultUnits>mg/Kg</ResultUnits>
        </ReportedResult>
</SamplePlusMethod>
```

3.2 RELATIONSHIPS

This section provides a discussion of the analytical relationships within the SEDD Specification. These relationships are used to link each reported sample result back to the underlying processes that were used to generate or evaluate that result.

3.2.1 Batches

- The SEDD Specification uses the concept of a batch as the primary mechanism for associating QC samples with regular ones. Analytical data review requires this association to assess the impact of QC sample performance to the quality of the regular sample results. For example, when doing volatiles by GC/MS, it is common practice to tune the instrument,

calibrate it, and analyze a blank once every shift. In this case, the SEDD Specification uses the concept of the Analysis batch to associate all the analyses done in one shift. All three QC analyses (tune, calibration, and blank) are associated with all regular sample analyses in that shift by having a common value for the AnalysisBatch data element that occurs in each of their Analysis nodes.

- The actual value used for the AnalysisBatch data element corresponding to one shift is not specified by the SEDD Specification, only that the value must be the same for all analyses in one shift and different for analyses in different shifts. The basis of the value for the AnalysisBatch data element must be given in the instructions provided by the data requester.

- The SEDD Specification uses the following eleven (11) data elements to define Batches:

  AnalysisBatch
  A group of analyses done on one instrument under the control of one continuing calibration or continuing calibration verification. Calibration is used in a generic sense. The details of what defines an AnalysisBatch depends on the method.

  PreparationBatch
  A group of aliquots prepared together for analysis by one method. 'Together' can imply similarity of the time, place, and manner of preparation, with details depending on the method. The notion of preparation is used in a generic sense for any activity prior to instrumental analysis. Method blanks and/or Laboratory Control Samples are often used to demonstrate that the laboratory's process is in control in each PreparationBatch.

  HandlingBatch
  A group of samples, not aliquots, handled together during the initial processing for analysis by one method. An example of QC in a HandlingBatch is a TCLP apparatus blank.

  CleanupBatch
  A group of aliquots going through a cleanup step together as part of preparation for analysis by one method. An example of QC in a CleanupBatch is a Gel Permeation Chromatography (GPC) calibration.

  RunBatch
  A group of analyses done on one instrument under the control of one initial calibration. Calibration is used in a generic sense. The details of what defines RunBatch depend on the method. Typically, one RunBatch includes the analyses from one or more analysis batches.

  MethodBatch
  A group of samples, not aliquots, with similar matrices, analyzed by one method and expected to have similar response to the method. Matrix spikes and duplicates are typical types of QC associated with a MethodBatch.

LabReportingBatch
A group of samples reported as a unit, (e.g., a CLP Sample Delivery Group). This batch is often used to define the context for definition (uniqueness) of batch values in data generated by the laboratory.

StorageBatch
A group of samples that are stored together. Volatile Organic Compound (VOC) refrigerator blanks are examples of QC associated with a StorageBatch.

ShippingBatch
A group of samples shipped in one container, such as a crate, cooler, or ice chest. Trip and temperature blanks are examples of QC associated with a ShippingBatch.

EquipmentBatch
A group of samples collected using the same equipment in a defined period of time. Rinsate blanks are examples of QC associated with an EquipmentBatch.

SamplingBatch
A group of samples collected together. Field blanks are examples of QC associated with a SamplingBatch.

The following example shows how a regular field sample is linked to the preparation, run, and analysis batches when implemented according to SW-846 Method 8270 (Semivolatile) rules. This example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- First tune: -->

<InstrumentQC>
        <QCType>Instrument_Performance_Check_Tune</QCType>
        <LabInstrumentQCID>Tune 1</LabInstrumentQCID>

        <Analysis>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>first analysis batch</AnalysisBatch>
        </Analysis>
</InstrumentQC>

<!-- First initial calibration: -->

<InstrumentQC>
        <QCType>Initial_Calibration</QCType>
        <LabInstrumentQCID>ICAL 1</LabInstrumentQCID>
```

```
        <Analysis>
            <AnalysisType>RRF-10</AnalysisType>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>first analysis batch</AnalysisBatch>
        </Analysis>
        <!-- and similarly for the RRF20, RRF50, RRF80 and RRF160 -->
</InstrumentQC>

<!-- Second tune performed at a later date: -->

<InstrumentQC>
        <QCType>Instrument_Performance_Check_Tune</QCType>
        <LabInstrumentQCID>Tune 2</LabInstrumentQCID>

        <Analysis>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>second analysis batch</AnalysisBatch>
        </Analysis>
</InstrumentQC>

<!-- First continuing calibration verification: -->

<InstrumentQC>
        <QCType>Continuing_Calibration_Verification</QCType>
        <LabInstrumentQCID>CCV1</LabInstrumentQCID>

        <Analysis>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>second analysis batch</AnalysisBatch>
        </Analysis>
</InstrumentQC>

<!-- Regular sample: -->

<SamplePlusMethod>
        <ClientSampleID>S001</ClientSampleID>

        <Analysis>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>second analysis batch</AnalysisBatch>

            <PreparationPlusCleanup>
                <PreparationBatch>first preparation batch</PreparationBatch>
            </PreparationPlusCleanup>
        </Analysis>
```

```
</SamplePlusMethod>

<!-- Matrix spike: -->

<SamplePlusMethod>
        <QCType>Matrix_Spike</QCType>

        <Analysis>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>second analysis batch</AnalysisBatch>

            <PreparationPlusCleanup>
                <PreparationBatch>first preparation batch</PreparationBatch>
            </PreparationPlusCleanup>
        </Analysis>
</SamplePlusMethod>

<!-- Matrix spike duplicate: -->

<SamplePlusMethod>
        <QCType>Matrix_Spike_Duplicate</QCType>

        <Analysis>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>second analysis batch</AnalysisBatch>

            <PreparationPlusCleanup>
                <PreparationBatch>first preparation batch</PreparationBatch>
            </PreparationPlusCleanup>
        </Analysis>
</SamplePlusMethod>

<!-- Method blank: -->

<SamplePlusMethod>
        <QCType>Method_Blank</QCType>

        <Analysis>
            <RunBatch>first run batch</RunBatch>
            <AnalysisBatch>second analysis batch</AnalysisBatch>

            <PreparationPlusCleanup>
                <PreparationBatch>first preparation batch</PreparationBatch>
            </PreparationPlusCleanup>
```

</Analysis>
</SamplePlusMethod>

As shown in the above example, the Client Sample ID "S001" is linked to three batches as follows:

- To the intial calibration and associated tune by the data element RunBatch containing the value "first run batch".
- To the continuing calibration verification and associated tune by the data element AnalysisBatch containing the value "second analysis batch".
- To the Method QC samples (Matrix Spike, Matrix Spike Duplicate, and Method Blank) by the data element PreparationBatch containing the value "first preparation batch".

3.2.2 Analysis Groups

- Some methods may require several analyses to calculate one result (e.g., Total Organic Carbon (TOC), method of standard additions). Further, some methods may allow several potential results to be calculated, with an average final result selected for reporting as the "final result." An example using the Method of Standard Additions is given below.

- The SEDD Specification uses AnalysisGroup nodes under the SamplePlusMethod node to associate several analyses underlying one set of possible results.

- AnalysisGroup nodes under the InstrumentQC node are also used to link several analyses (e.g., multipoint initial calibrations, however, single-point initial calibrations can also be reported in this manner) that are part of one instrument QC process.

- The AnalysisGroupID data element must be present in the following three nodes - AnalysisGroup, Analysis, and ReportedResult.

The following example, derived from the SW-846 Inorganics data, illustrates a complex method requiring AnalysisGroup nodes. Data is shown for only one analyte and many less important data elements are omitted. In this example, the method of standard additions (MSA) was performed on the field soil sample "BH-0945Q". This example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- Regular soil sample - GFAA Method with MSA data included: -->

<SamplePlusMethod>
        <!-- identification -->
        <ClientMethodID>7841</ClientMethodID>
        <ClientSampleID>BH-0945Q</ClientSampleID>
        <LabSampleID>980917-011</LabSampleID>
        <MatrixID>Soil</MatrixID>

        <!-- Final result: -->

        <ReportedResult>
            <!-- linkage to the AnalysisGroup for this reported result -->
            <AnalysisGroupID>First Analysis Group</AnalysisGroupID>
            <!-- identification -->
            <ClientAnalyteID>7440-28-0</ClientAnalyteID>
            <AnalyteName>Thallium</AnalyteName>
            <!-- Final Result and associated units -->
            <LabQualifiers/>
            <Result>1.58</Result>
            <ResultUnits>mg/kg</ResultUnits>
        </ReportedResult>

        <!-- ANALYSIS GROUP for MSA -->

        <AnalysisGroup>
            <AnalysisGroupID>First Analysis Group</AnalysisGroupID>
            <AnalysisType>MSA</AnalysisType>
        </AnalysisGroup>

        <!-- First MSA analysis -->

        <Analysis>
            <AnalysisGroupID>First Analysis Group</AnalysisGroupID>
            <LabAnalysisID>GFAA-TL1-981015-28</LabAnalysisID>
            <AnalysisType>MSA-0</AnalysisType>

            <Analyte>
                <ClientAnalyteID>7440-28-0</ClientAnalyteID>
                <IntermediateResult>7.6</IntermediateResult>
                <IntermediateResultUnits>ug/L</IntermediateResultUnits>
            </Analyte>
        </Analysis>

        <!-- Second MSA analysis: -->
```

```
        <Analysis>
            <AnalysisGroupID>First Analysis Group</AnalysisGroupID>
            <LabAnalysisID>GFAA-TL1-981015-29</LabAnalysisID>
            <AnalysisType>MSA-1</AnalysisType>

            <Analyte>
                <ClientAnalyteID>7440-28-0</ClientAnalyteID>
                <ExpectedResult>3.5</ExpectedResult>
                <IntermediateResult>13</IntermediateResult>
                <IntermediateResultUnits>ug/L</IntermediateResultUnits>
            </Analyte>
        </Analysis>

        <!-- Third MSA analysis: -->

        <Analysis>
            <AnalysisGroupID>First Analysis Group</AnalysisGroupID>
            <LabAnalysisID>GFAA-TL1-981015-30</LabAnalysisID>
            <AnalysisType>MSA-2</AnalysisType>

            <Analyte>
                <ClientAnalyteID>7440-28-0</ClientAnalyteID>
                <ExpectedResult>7.0</ExpectedResult>
                <IntermediateResult>17</IntermediateResult>
                <IntermediateResultUnits>ug/L</IntermediateResultUnits>
            </Analyte>
        </Analysis>

        <!-- Fourth MSA analysis: -->

            <Analysis>
                <AnalysisGroupID>First Analysis Group</AnalysisGroupID>
                <LabAnalysisID>GFAA-TL1-981015-31</LabAnalysisID>
                <AnalysisType>MSA-3</AnalysisType>

                <Analyte>
                    <ClientAnalyteID>7440-28-0</ClientAnalyteID>
                    <ExpectedResult>10.5</ExpectedResult>
                    <IntermediateResult>21</IntermediateResult>
                    <IntermediateResultUnits>ug/L</IntermediateResultUnits>
                </Analyte>
            </Analysis>
    </SamplePlusMethod>
```

As shown in the above example, the Client Sample ID "BH-0945Q" has been analyzed using the Method of Standard Additions to determine its final Thallium Result "1.58" with ResultUnits "mg/kg". The final reported result is linked to the data element AnalysisGroupID with a value of "First Analysis Group" instead of to a single analysis (which would have used the data element LabAnalysisID). All the MSA Analyses are similarly linked to each other and the final result through the same AnalysisGroupID data element having the value "First Analysis Group".

3.2.3 Analyte Groups

- Some methods may require the reporting of an analyte that is not directly measured by that method where the reported analyte is actually the combination of two or more analytes that are directly measured by that method (e.g., Hardness, where the reported analyte Hardness is actually the summed values of the Calcium and Magnesium measured analytes).

- The SEDD Specification uses AnalyteGroup nodes under the Analysis node to associate two or more analytes that are directly measured using the method(s) indicated to report an analyte that typically cannot be directly measured.

- The AnalyteGroupID data element must be present in the following three nodes - AnalyteGroup, Analyte, and ReportedResult.

The following example, derived from SW-846 Inorganics data, illustrates a method requiring the AnalyteGroup node. Data is shown for only two measured analytes and many less important data elements are omitted. In this example, a single analysis was performed on the field water sample "SW-001" and a final result was reported for 'Hardness'. This example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

<!-- *Regular water sample - ICP Method:* -->

<SamplePlusMethod>
<!-- *identification* -->
<ClientMethodID>6010B</ClientMethodID>
<ClientSampleID>SW-001</ClientSampleID>
<LabSampleID>040817-010</LabSampleID>
<MatrixID>Water</MatrixID>

<!-- *Final result:* -->

<ReportedResult>
<!-- *linkage to the AnalyteGroup for this reported result* -->
<AnalyteGroupID>First Analyte Group</AnalyteGroupID>
!-- *identification* -->
<ClientAnalyteID>Hardness</ClientAnalyteID>
<ClientAnalyteName>Hardness</ClientAnalyteName>

```
            <!-- Final Result and associated units -->
            <LabQualifiers/>
            <Result>28</Result>
            <ResultUnits>mg/L</ResultUnits>
        </ReportedResult>

        <Analysis>
            <LabAnalysisID>ICP-981015-28</LabAnalysisID>
            <AnalysisType>Initial</AnalysisType>

            <!-- ANALYTE GROUP for Hardness -->

            <AnalyteGroup>
                <AnalyteGroupID>First Analyte Group</AnalyteGroupID>
                <!-- identification -->
                <ClientAnalyteID>Hardness</ClientAnalyteID>
                <ClientAnalyteName>Hardness</ClientAnalyteName>
            </AnalyteGroup>

        <!-- First Analyte to be associated with Hardness -->

            <Analyte>
                <AnalyteGroupID>First Analyte Group</AnalyteGroupID>
                <!-- identification -->
                <ClientAnalyteName>Calcium</ClientAnalyteName>
                <ClientAnalyteID>7440-70-2</ClientAnalyteID>
                <!-- Final Result and associated units -->
                <Result>15</Result>
                <ResultUnits>mg/L</ResultUnits>
            </Analyte>

            <!-- Second Analyte to be associated with Hardness -->

            <Analyte>
                <AnalyteGroupID>First Analyte Group</AnalyteGroupID>
                <!-- identification -->
                <ClientAnalyteName>Magnesium</ClientAnalyteName>
                <ClientAnalyteID>7439-95-4</ClientAnalyteID>
                <!-- Final Result and associated units -->
                <Result>13</Result>
                <ResultUnits>mg/L</ResultUnits>
            </Analyte>
        </Analysis>
</SamplePlusMethod>
```

As shown in the above example, the Client Sample ID "SW-001" has been analyzed to determine its final Hardness Result "28" with ResultUnits "mg/L". The final reported result is linked to the data element AnalyteGroupID with a value of "First Analyte Group" instead of to a single Analysis/Analyte (which would have used the data element LabAnalysisID). All the Analytes that are associated with the final reported Hardness result are similarly linked to each other and the final result through this same data element AnalyteGroupID having the value "First Analyte Group".

3.2.4 QC Categories and QC Linkage

- The QCCategory and QCLinkage data elements (under the SamplePlusMethod and InstrumentQC nodes - see Figures 10 and Figure 11) support cross-program movement of data (e.g., supply data for processing by different site's data validation programs) and development of implementation-independent EDD processing software. Because valid values for both the QCCategory and QCLinkage data elements are given by the SEDD Specification, processing software can 'understand' how to store, report, and associate a method-level QC sample.

- The QCCategory data element specifies the type of QC values and comparisons appropriate to a method-level QC sample (one reportable in a SamplePlusMethod node). Valid values for the QCCategory data element include:

  Blank
  A QC sample made to contain negligible (or unmeasurable) quantities of the analyte(s) of interest. Detection of an analyte in a blank is an indication of contamination. Data is similar to that for a regular sample. This same QCCategory would be used with many QCTypes (e.g., those corresponding to method, trip, rinsate, field, and other types of implementation-defined blanks).

  Blank_Spike
  A QC sample with known amounts of target analytes added to a Blank. Data potentially includes a PercentRecovery and an ExpectedResult for each analyte for which the AnalyteType is Spike. An Interference Check Sample is a type of BlankSpike where the results for unspiked analytes are of interest.

  Spike
  A reanalysis of a regular sample done for QC purposes with known amounts of target analytes added to the sample. Data is expected to include an OriginalClientSampleID for each SamplePlusMethod node and, potentially, a PercentRecovery and an ExpectedResult for each analyte for which the AnalyteType is Spike. The OriginalClientSampleID identifies the original sample that was spiked.

Duplicate
A reanalysis of a regular sample done for QC purposes. Data is expected to include an OriginalClientSampleID for each SamplePlusMethod node and, potentially, an RPD for each analyte. The OriginalClientSampleID identifies the original sample that was reanalyzed.

This use of Duplicate is meant to include what are called 'splits' or 'replicates', as long as only two sets of results are involved.

Serial_Dilution
A diluted reanalysis of a regular sample done for QC purposes. Data is expected to include an OriginalClientSampleID and, potentially, a PercentDifference. The OriginalClientSampleID identifies the original sample that was reanalyzed.

Blank_Spike_Duplicate
A second analysis of a BlankSpike. Data is expected to include an OriginalLabSampleID, and, potentially, a PercentRecovery, an RPD and an ExpectedResult for each analyte for which the AnalyteType is 'Spike'. The OriginalLabSampleID identifies the original Blank_Spike that was reanalyzed.

Spike_Duplicate
A second reanalysis of a regular sample done for QC purposes with a known spike added. Data is expected to include an OriginalClientSampleID, and, potentially, a PercentRecovery, an RPD and an ExpectedResult for each analyte where the AnalyteType is 'Spike'. The OriginalClientSampleID identifies the regular sample that was spiked and reanalyzed. There should be another SamplePlusMethod node with QCCategory 'Spike' with the same ClientMethodID and OriginalClientSampleID.

Non-Client_Sample
This type of QC sample is different than the other QC sample types normally encountered in that it is, simply, a regular sample. Many programs require that certain data be reported from the analysis of non-client samples that were prepared or analyzed with project-specific samples. This is often done to verify run sequences and to ensure that certain QC samples were analyzed at an appropriate frequency. The actual analytical results would seldom be reported.

The QCLinkage data element specifies the type of batch used to make associations for a QC sample and the following applies:

- Allowed valid values for the QCLinkage data element are the names of the various batch fields defined by the SEDD Specification. They are AnalysisBatch, PreparationBatch, HandlingBatch, CleanupBatch, RunBatch, MethodBatch, LabReportingBatch, StorageBatch, ShippingBatch, EquipmentBatch and SamplingBatch (as given in Section 3.2.1). Because there are potentially multiple handling nodes for each sample, if the QCLinkage data element value is HandlingBatch, a ClientMethodID or HandlingType data element must be specified for the QC sample to make the linkage unambiguous. Similarly, a ClientMethodID or

CleanupType data element must be specified if the QCLinkage data element value is CleanupBatch.

- The limited list of valid values for QCCategory and QCLinkage data elements are part of the SEDD Specification. They represent a choice about the types of data expected to be reported. This choice enables the development of reasonably method-independent database software to support performance-based verification and validation of analytical data.

As an example, in a SamplePlusMethod node, if the QCType data element value is Lab_Duplicate, the QCCategory data element value is Duplicate, and the QCLinkage data element value is MethodBatch, then a reader will know this data is for a client-defined type of QC called a Lab_Duplicate. The reader will also know it is processed with rules typical for Duplicates and it is to be associated with other SamplePlusMethod nodes with the same value for the MethodBatch data element. The following example shows how the QCCategory and QCLinkage data elements in the EDD help to show the type of sample being analyzed (e.g., Blank) and the linkage to a batch (e.g., Preparation) in an Inorganics analysis:

```
<SamplePlusMethod>
    <ClientMethodID>6010C</ClientMethodID>
    <LabSampleID>ICP-WG2413-1</LabSampleID>
    <MatrixID>Soil</MatrixID>
    <QCCategory>Blank</QCCategory>
    <QCLinkage>PreparationBatch</QCLinkage>
    <QCType>Method_Blank</QCType>

    <Analysis>
        <PreparationPlusCleanup>
            <PreparationBatch>ICP-WG2413-A</PreparationBatch>
        </PreparationPlusCleanup>
    </Analysis>
</SamplePlusMethod>
```

A general EDD reader might not recognize the ClientMethodID data element value (6010C, specified by a specific implementation) and hence not know the meaning of the QCType data element value (Method_Blank) or even that this is ICP inorganics data. However, based on the QCCategory data element value (Blank) it knows to use “blank-like” rules to process this data. Based on the QCLinkage data element value (PreparationBatch) it knows to associate this QC sample with other analyses with the same value (ICP-WG2413-A) for the PreparationBatch data element.

3.2.5 Comparisons

There are two basic types of comparisons that are allowed in the SEDD Specification - Analyte Comparisons and Peak Comparisons.

Analyte Comparison

- Analyte comparisons are used to describe the effects of potentially interfering analytes on the measured analyte peak.

- Data from instrument QC checks for cross-analyte interference are reported in the AnalyteComparison node.

- A common use of AnalyteComparison node is in the reporting of ICP Interelement Correction factors where the contribution to the analyte measurement from interfering analytes is reported in data elements in AnalyteComparison nodes under a Peak node (see Figure 7). This is shown in the following XML example file that starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- ICP Interelement Correction Factors: -->

<InstrumentQC>
    <ClientMethodID>6010C</ClientMethodID>
    <LabInstrumentQCID>IEC 1</LabInstrumentQCID>

    <Analysis>
        <LabAnalysisID>P2_ICP_042694</LabAnalysisID>

        <!-- Report interference with this analyte: -->

        <Analyte>
            <ClientAnalyteID>7439-92-1</ClientAnalyteID>
            <AnalyteName>Lead</AnalyteName>

            <!-- At this peak: -->

            <Peak>
                <!-- identified by integer wavelength -->
                <PeakID>220</PeakID>
                <!-- described by actual value -->
                <Wavelength>220.35</Wavelength>

            <!-- First interferent: -->
                <AnalyteComparison>
                    <!-- identified by CAS number -->
```

```
                    <ClientAnalyteID>7429-90-5</ClientAnalyteID>
                    <AnalyteName>Aluminum</AnalyteName>
                    <CorrectionFactor>0.000370</CorrectionFactor>
                </AnalyteComparison>
            </Peak>
        </Analyte>
    </Analysis>
</InstrumentQC>
```

As shown in the above example, during the analysis of Lead by ICP-AES, Aluminum is an interferent. Lead is identified as an analyte under the Analyte node using the data element AnalyteName. The Peak used for measurement of Lead is identified in the Peak node by the data element PeakID. Aluminum interferes with this peak and its interelement correction factor is reported in the AnalyteComparison node in the data element CorrectionFactor.

Peak Comparison

- Peak comparisons are used to compare measurements made at two or more different peaks. Peak comparisons can describe cross-peak comparisons within the same analyte (e.g., abundance ratios for tunes), between two analytes (e.g., calculating the relative response factor (RRF) for initial and continuing calibrations or inter-peak resolutions in chromatography).

- Data from these peak comparisons are reported in the PeakComparison node (see Figures 6 and 7).

One common example for using PeakComparison nodes (given below) is for the reporting of GC/MS tune data where only a single analyte is involved. The following example XML file starts with a comment line. Please note that there are comment lines in XML format to clarify information presented in the nodes or data elements following the comment lines.

```
<!-- Part of the data for one tune: -->

    <InstrumentQC>
        <ClientMethodID>8270C</ClientMethodID>
        <QCType>Instrument_Performance_Check_Tune</QCType>
        <LabInstrumentQCID>Tune 1</LabInstrumentQCID>

    <Analysis>
        <LabAnalysisID>SV417</LabAnalysisID>

        <Analyte>
            <ClientAnalyteID>5074-71-5</ClientAnalyteID>
            <ClientAnalyteName>DFTPP</ClientAnalyteName>
            <AnalyteType>System_Monitoring_Compound</AnalyteType>
```

```
            <!-- One mass spectral peak: -->

            <Peak>
                <!-- identified by nominal mass -->
                <PeakID>68</PeakID>

                <!-- First comparison: -->

                <PeakComparison>
                    <PeakID>198</PeakID>
                    <PercentRatio>0.0</PercentRatio>
                </PeakComparison>

                <!-- Second Comparison: -->

                <PeakComparison>
                    <PeakID>69</PeakID>
                    <PercentRatio>0.0</PercentRatio>
                </PeakComparison>
            </Peak>
        </Analyte>
    </Analysis>
</InstrumentQC>
```

As shown in the above example, a DFTPP tune (for method SW-846 8270 semivolatile organics) is being reported under the Analyte node using the data element AnalyteName. The peak (mass number 68) to be evaluated is identified under the Peak node using the data element PeakID. This peak (mass number 68) is compared to the following two peaks (mass numbers 198 and 69) under separate PeakComparison nodes using the data element PeakID. The data element PercentRatio is used to report the actual comparison (expressed as a percent ratio).