

Water Quality Exchange XML Training Manual for BEACH Monitoring Data

U.S. Environmental Protection Agency
Washington, DC
September 18, 2008

EPA Technical Support
1-800-424-9067
storet@epa.gov

Electronic copies of this document, along with many others,
are available at <http://www.epa.gov/storet>

[This page intentionally left blank.]

Table of Contents

Definitions.....	v
Introduction.....	1
Overview of BEACH Data Submission Process Changes.....	1
Structure of BEACH WQX Data Submissions.....	2
Submission File.....	2
Header Document	2
Header	3
Payload.....	3
Payload Data Structure at a Glance	4
Data Elements	5
Organization.....	5
Project	5
Monitoring Location	6
Activity	7
Activity Description.....	8
Sample Description.....	9
Result	10
Result Description.....	10
Result Analytical Method	11
Result Lab Information	12
Trip Blank Activity	12
Activity Group	13
Updating and Inserting Data into WQX	14
Insert Process	14
Update Process.....	14
Deleting Data from WQX.....	15
Delete Process.....	15
Summary	16
Appendix A.....	19

List of Figures

1. Structure of a Header Document.....	3
2. Overview of WQX XML schema structure for this tutorial.	4
3. <Project> data elements.....	6
4. <MonitoringLocation> data elements.....	7
5. <Activity> data elements.....	8
6. <ActivityDescription> data elements.	8
7. More detailed <ActivityDescription> data elements.	9
8. <SampleDescription> data elements.	10
9. <Result> data elements.....	10
10. <ResultDescription> data elements.	11
11. <ResultDescription> data elements for non-detect.....	11

List of Figures (continued)

12. <ResultAnalyticalMethod> data elements..... 12
13. <ResultLabInformation> data elements..... 12
14. Example trip blank activity..... 13
15. <ActivityGroup> data elements..... 14
16. Delete XML file..... 16

List of Tables

1. Comparison of the Old eBeaches Approach and the New WQX Approach to Submitting
BEACH Monitoring Data. 2
2. Elements Used in Header of WQX Submission. 3
3. General Rules that Apply to Updates of BEACH Monitoring Data..... 15

Definitions

Block—A group of data elements in an Extensible Markup Language (XML) file. A block includes everything between a matching pair of open and close tags. For example, the electronic address block could include the following two data elements: electronic address text and electronic address type.

Central Data Exchange (CDX)—The U.S. Environmental Protection Agency's (EPA's) electronic reporting mechanism that is used to submit legally acceptable data for various programs into EPA databases.

Data Element—A data field, such as Project Name.

Extensible Markup Language (XML)—A format for transferring data by including a self describing tag on each data element that identifies the data element. An example of a self describing tag is <ProjectName>. XML files are text documents that include tags for each data element.

Header—The portion of the submission file that contains basic information about the data submission.

Header Document—The .xml portion of a submission file that contains one header and one payload. The header document is the only portion of a submission file that is allowed to have an .xml extension.

Payload—The portion of the submission file that contains the Water Quality Exchange (WQX) monitoring data being submitted.

Primary Component—A set of data elements or group of data elements in an XML file that must be uniquely identified. A primary component includes everything between a matching pair of open and close tags. In WQX, the four primary components are <Project>, <Monitoring Location>, <Activity>, and <Activity Group>. The WQX XML schema requires at least one of these primary components to be included.

Reusable Data Block—A repeatable group of data elements in an XML file. For example, a depth or height measure element will always have a measurement value and a measurement unit, so these are repeatable elements.

Root Component—The highest level of data elements in an XML file. The root component is a set of XML tags that contain all other data elements in the schema. In WQX, the root component is Organization.

Schema—A defined, customized group of data elements that is organized in a set structure. EPA has defined an XML schema for WQX.

Submission File—A ZIP file containing an entire WQX data submission, which must consist of a header document along with optional attached binary objects.

Tag—A pair of identifiers used in XML files to label specific data elements. For example, a project name of “Chesapeake Bay” would be identified by the tag `<ProjectName>Chesapeake Bay</ProjectName>`. An open tag has no /, and a close tag starts with a /.

Water Quality Exchange (WQX)—A data flow (or data transfer process) that uses an XML schema to enable states, Tribes, EPA, and others to share water quality monitoring data.

Introduction

Welcome to the Beaches Environmental Assessment and Coastal Health (BEACH) Program's Water Quality Exchange (WQX) Training Manual. WQX is a data flow (or data transfer process) that uses an Extensible Markup Language (XML) schema to enable you to load data into the U.S. Environmental Protection Agency's (EPA's) Storage and Retrieval (STORET) warehouse using XML files. XML allows you to format the data being imported and ensures that the data are consistent with the WQX XML schema. Once proper XML files are created that are consistent with the schema, you can migrate BEACH monitoring data into the STORET warehouse.

This document will help you become familiar with the formatting protocol for submitting BEACH monitoring data using the WQX XML schema and how to properly use WQX data elements. This training manual covers the required data elements for an XML file based on the WQX schema, but additional fields may apply in special cases. By the end of the tutorial, you should be able to

- Identify WQX XML blocks and tags.
- Understand and be able to apply basic WQX business rules and data structure to BEACH monitoring data.
- Create a BEACH monitoring data file to insert, update, or delete from the STORET warehouse using the WQX data flow.
- Understand what activity groups and trip blanks are and how they can be documented using XML.

This training manual includes both "WQX_BEACH_Demotest.xml" and "WQX_BEACH_Demotest_Delete.xml" files, which contain sample data that may be used to assist you in working through this tutorial. Please take a moment to find them on your computer using Microsoft Windows Explorer or a similar navigation program. You can open these files in a browser, such as Internet Explorer, or a text editor, such as Microsoft WordPad. Because you will be accessing the "WQX_BEACH_Demotest.xml" and "WQX_BEACH_Demotest_Delete.xml" in this tutorial, you should write down the full path location on a piece of paper.

Overview of BEACH Data Submission Process Changes

Because WQX relies entirely on XML for data transfer, the BEACH data submission process has changed to accommodate this difference. This transfer of data to the STORET warehouse no longer involves the use of Web Registration or WebSIM. All data being transferred must be included in an XML file. Table 1 outlines the differences between the former electronic Beach Environmental Assessment and Coastal Health System (eBeaches) approach to BEACH data submission and the WQX approach to data submission.

Table 1. Comparison of the Old eBeaches Approach and the New WQX Approach to Submitting BEACH Monitoring Data.

eBeaches	WQX
Submitting Projects	
Web registration through CDX	XML file is submitted through CDX, then WQX inserts, updates, or deletes projects
Submitting Stations	
Web registration through CDX or flat file submitted using WebSIM through CDX	XML file is submitted through CDX, then WQX inserts, updates, or deletes stations
Submitting Monitoring Results	
XML file is submitted using WebSIM through CDX	XML file is submitted through CDX, then WQX inserts, updates, or deletes results
Login Process	
CDX username and password plus database (WATERS9i) username and password	Originating Node Username and Password passed to CDX. CDX authentication returns security token
XML Validation Process	
Upload of XML BEACH monitoring data to CDX validated against schema. Returns e-mail to CDX inbox, confirming status of validation	XML submission is validated against WQX schema within CDX. Validation Report is available at CDX for download
Error Reporting	
STORET errors are provided by WebSIM in Results Import Log	E-mail notification of "Completed" or "Failed" status is sent to the submitter. WQX errors are provided in a processing report available at CDX for download

Structure of BEACH WQX Data Submissions

To properly submit, update, or delete BEACH monitoring data to EPA's STORET data warehouse, the data must meet certain CDX and WQX structural and processing requirements. All data must be submitted in XML format, have header information attached, and be contained in a header document. The header document is zipped up in a ZIP file with any attached binary objects (.pdf or .jpg) to create a submission file.

Submission File

The submission of BEACH monitoring data through CDX must consist of a submission file, which is a ZIP file that contains a header document and any attached binary objects. The purpose of creating a ZIP file is to not only consolidate the different portions of the submission into one file, but also to increase the efficiency of the transmission due to the decrease in file size. The maximum allowable size of a submission file is 80 megabytes. The header document is the only file within a submission that is allowed to have a .xml extension.

Header Document

A submission must consist of at least a header document, but may also contain other attached binary files. The header document, which is the only file allowed to have a .xml extension, must

consist of one header and one payload. The header document will always begin with a tag, `<?xml version="1.0" encoding="UTF-8" ?>`, followed by a `<Document>` tag that determines your document ID and the path referencing the schema with which to validate your submission. The header document must end with the closing tag of `</Document>`. Figure 1 shows the structure of a header document.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <Document Id="UI34" xmlns="http://www.exchangenetwork.net/schema/v1.0/ExchangeNetworkDocument.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
+ <Header>
+ <Payload Operation="Update-Insert">
</Document>

```

Figure 1. Structure of a Header Document

Header

The header section of the header document provides basic information about the data submission. The header can be inserted into an XML document containing the BEACH payload by either copying and pasting it from another BEACH WQX document or by using a header generation tool that is available from <http://test.epacdxnode.net/helpdesk>. Table 2 shows the elements used in the WQX header section.

Table 2. Elements Used in Header of WQX Submission.

Element	Description	Example Value	Required
Author	First and last name of the individual generating the XML document	Joe Smith	Yes
Organization	Name of company, agency, or individual generating the XML document	State X Department of Environmental Quality	Yes
Title	Type of submission	Must be "WQX"	Yes
Creation Time	Date/time when the document was generated	2003-01-01T12:12:12 (Where date is a valid XML date format string)	Yes
Comment	Free text description of the message contents		No
Contact Information	Name, mailing address, city, state, zip code, telephone number, and e-mail address of person who may be contacted with questions about the submission	Joe Smith 123 Main Street Portland, OR 97226 503-123-4567 Joe@deq.statex.gov	Yes

Payload

The payload portion of a data submission contains the WQX data that you are sending in XML format. There are two different formats that the payload must conform to depending on the purpose of your submission. If you are attempting to add data to the warehouse or modify data that is already in the warehouse, you will use the Update-Insert schema. If you are attempting to delete data from the warehouse, you will use the Delete schema. The payload section must

reference the appropriate schema as either WQX_WQX_v1.0.xsd or WQX_WQX_Delete_v1.0.xsd.

The payload contains an attribute called “Operation,” which determines whether your submission is an “Update-Insert” or a “Delete.” In XML, the following format appears: <Payload Operation=“Update-Insert”> or <Payload Operation=“Delete”>. The end of the payload section must contain a closing tag of </Payload>. All of your monitoring data will be contained between these tags. Specific monitoring data format for WQX will be discussed later in this tutorial.

Payload Data Structure at a Glance

To understand how WQX transmits your data in an XML file, it is important that you understand how the data are structured. The WQX data structure follows the way organizations generally report water quality monitoring.

- An **organization** is the group that collects water quality monitoring information.
- An organization may have one or more **projects** to collect data.
- **Monitoring locations** are places where monitoring activities occur.
- **Monitoring activities** are conducted by an organization’s staff.
- **Results** are the data collected or produced from monitoring activities.
- **Activity groups** are categories of activities, such as field sets, replicates, subsamples, and quality control (QC) samples.

In the demonstration WQX file we will be working with (see Figure 2), as with all WQX XML files, all of the data elements are included within the Organization group or root component. For this tutorial, the organization is called 21RIBCH_WQX. Figure 2 illustrates how WQX is structured to follow monitoring steps. If you count the tags of the different primary components, you will see that we will be working with the data elements for one project, one monitoring location, two activities, and one activity group for this organization.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <Document Id="UI34" xmlns="http://www.exchangenetwork.net/schema/v1.0/ExchangeNetworkDocument.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
+ <Header>
- <Payload Operation="Update-Insert">
  - <WQX xmlns="http://www.exchangenetwork.net/schema/wqx/1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
    instance" xsi:schemaLocation="http://www.exchangenetwork.net/schema/wqx/1
    http://www.exchangenetwork.net/schema/wqx/1/0/WQX_WQX_v1.0.xsd">
    - <Organization>
      + <OrganizationDescription>
      + <ElectronicAddress>
      + <Telephonic>
      + <OrganizationAddress>
      + <OrganizationAddress>
      + <Project>
      + <MonitoringLocation>
      + <Activity>
      + <Activity>
      + <ActivityGroup>
      </Organization>
    </WQX>
  </Payload>
</Document>

```

Figure 2. Overview of WQX XML schema structure for this tutorial.

Data Elements

If you review an XML file with a browser, such as Internet Explorer, you can click on the “+” sign next to the tag, which will provide more in-depth information. We will explore the following components in this tutorial:

- Organization
- Project
- Monitoring location
- Activity
- Result
- Activity group.

Organization

The BEACH program, similar to other programs, requires that all organizations performing monitoring activities be issued a unique identifier to differentiate it from other organizations within the STORET warehouse. In the case of the BEACH program, this identifier is given to the state agency performing the monitoring. The <OrganizationIdentifier> must be used every time an organization wants to insert data into, update, or delete data from the STORET warehouse using the WQX data flow.

All of the information in an XML file is contained within an organization. Within the Organization root component, there are several primary data elements, including organization address, telephone information, electronic address, organization address, project, monitoring location, and activity.

The Organization root component has only a few simple business rules, as indicated in the *Water Quality Exchange Flow Configuration Document*¹, which defines WQX data services, approaches and processes used to exchange information, and the *Water Quality Exchange Data Dictionary*², which defines all WQX data elements and associated business rules. For example, the Organization Identifier and the Formal Name of the Organization fields are required. No other fields are required, although a few are conditionally required if certain other information is provided. For example, if a Telephone Number is provided for the organization, then the Telephone Number Type is required.

Project

Within the BEACH program, a project refers to an individual beach to be monitored. In other words, the <ProjectIdentifier> is really a beach identifier. Figure 3 shows the required information for projects in the WQX XML schema. Within the <Project> primary component, the data are stored between tags of related fields. For example,

¹ U.S. EPA (Environmental Protection Agency). 2006. *Water Quality Exchange Flow Configuration Document, Version 1.0*. U.S. EPA, Office of Water, Washington, DC. November 15. Available at <http://www.exchangenetwork.net/exchanges/water/wqx.htm>.

² U.S. EPA (Environmental Protection Agency). 2006. *Water Quality Exchange Data Dictionary*. U.S. EPA, Office of Water, Washington, DC. September 13. Available at <http://www.exchangenetwork.net/exchanges/water/wqx.htm>.

- The unique project ID is RIBEACH-001, and the data are labeled with the <ProjectIdentifier> open and close tags.
- The project name is labeled by the <ProjectName> tags.
- The project description is labeled with the <ProjectDescriptionText> tags.

If you use a text editor to view an XML file (as shown in Figure 3), you will see the corresponding data information for each field between the open and close tags.

```
+ <OrganizationAddress>
- <Project>
  <ProjectIdentifier>RIBEACH-001</ProjectIdentifier>
  <ProjectName>Sandy Beach</ProjectName>
  <ProjectDescriptionText>BEACH Act BeachID for linking to water quality sample</ProjectDescriptionText>
</Project>
+ <MonitoringLocation>
+ <Activity>
```

Figure 3. <Project> data elements.

The <Project> primary component has two required fields and one optional field. The two required fields are Project Identifier and Project Name. The Project Identifier, which can be up to 35 characters long, can be reused to update or edit project data in WQX without repeating all of the data fields in XML. Project Description is an optional field, but it is recommended so that a user of the data can understand what the project is about.

The <Project> primary component may also contain a binary large object, such as a PDF, that contains additional project information. The fields containing the file name for the binary large object and the text extension of the file are only required if the <AttachedBinaryObject> header is inserted.

Monitoring Location

The <MonitoringLocation> primary component represents the stations used for monitoring at your beach (project). As mentioned in the Project section of this tutorial, each project in the BEACH program represents an individual beach. Each individual beach can be monitored in multiple locations, which would be reported separately.

Next, expand <MonitoringLocation> and further expand <MonitoringLocationIdentity> and <MonitoringLocationGeoSpatial> to view the data between the tags for the fields of monitoring locations. As shown in Figure 4, there is only one location for this beach, ID RIBEACH-001-1, which is located at the western end of Sandy Beach. Its <MonitoringLocationTypeName> is BEACH Program Site-Ocean. <MonitoringLocationTypeName> is the equivalent of what Primary Station Type was in STORET, but unlike STORET, WQX no longer breaks out a Secondary Station Type. The business rules for monitoring location require that the following information is provided: an ID, a name, the type of location, and the geospatial information.

```

+ <Project>
- <MonitoringLocation>
  - <MonitoringLocationIdentity>
    <MonitoringLocationIdentifier>RIBEACH-001-1</MonitoringLocationIdentifier>
    <MonitoringLocationName>Sandy Beach 1</MonitoringLocationName>
    <MonitoringLocationTypeName>BEACH Program Site-Ocean</MonitoringLocationTypeName>
    <MonitoringLocationDescriptionText>This station is located at the western end of Sandy
      Beach.</MonitoringLocationDescriptionText>
  </MonitoringLocationIdentity>
  - <MonitoringLocationGeospatial>
    <LatitudeMeasure>41.32</LatitudeMeasure>
    <LongitudeMeasure>-71.81</LongitudeMeasure>
    <SourceMapScaleNumeric>100000</SourceMapScaleNumeric>
    <HorizontalCollectionMethodName>Interpolation-Map</HorizontalCollectionMethodName>
    <HorizontalCoordinateReferenceSystemDatumName>NAD83</HorizontalCoordinateReferenceSystemDatumName>
  </MonitoringLocationGeospatial>
</MonitoringLocation>
+ <Activity>

```

Figure 4. <MonitoringLocation> data elements.

Please note that some tags must contain a value from a limited list of options. For example, BEACH Program Site-Ocean is one of a specified set of allowed values for <MonitoringLocationTypeName>. EPA has created *Water Quality Exchange Domain Value Lists*³ to show which tags require specific choices and what the allowable values are. Appendix A contains a list of tags that have allowable values.

In the <MonitoringLocationGeospatial> block, we can see that the location is geospatially located at latitude 41.32 degrees, and longitude -71.81 degrees by the “Interpolation-Map” method using the “NAD83” datum. As the business rules in the *Water Quality Exchange Data Dictionary* indicate, “Interpolation-Map” is the only Horizontal Collection Method that requires a Source Map Scale Numeric value to be entered. Figure 4 shows that the scale of the map used to determine the latitude and longitude was 1:100,000 by placing the integer “100000” between the <SourceMapScaleNumeric> tags.

In addition to the <HorizontalCollectionMethodName>, which is a required, reusable data block, a <VerticalMeasure> reusable data block may be inserted in the <MonitoringLocationGeospatial> block. The <VerticalMeasure> block is optional, but if used must contain a vertical measure value, the vertical measure units, the vertical measure collection name, and the vertical coordinate reference system datum name. The <HorizontalCollectionMethodName> and <VerticalMeasure> blocks are the first examples of reusable data blocks in this tutorial.

Activity

Each separate activity in the <Activity> primary component represents a unique sampling or measurement event. Expand one of the three <Activity> blocks by clicking the “+” sign. As shown in Figure 5, this reveals in-depth information about the <ActivityDescription>, <ActivityLocation>, <SampleDescription>, and two <Results> blocks. Of these four block types, only <ActivityDescription> is required.

³ U.S. EPA (Environmental Protection Agency). 2006. *Water Quality Exchange Domain Value Lists*. U.S. EPA, Office of Water, Washington, DC. September 13. Available at http://www.epa.gov/STORET/future_downloads.html.

```

+ <MonitoringLocation>
- <Activity>
  + <ActivityDescription>
  + <SampleDescription>
  + <Result>
  + <Result>
  </Activity>
+ <Activity>
+ <ActivityGroup>

```

Figure 5. <Activity> data elements.

Activity Description

In the <ActivityDescription> block, the project and monitoring locations are assigned to the activity. As shown in Figure 6, we see that the unique activity, 01-23-456-78, has been conducted for the project RIBEACH-001 and at the location RIBEACH-001-1.

```

- <Activity>
  - <ActivityDescription>
    <ActivityIdentifier>01-23-456-78</ActivityIdentifier>
    <ActivityTypeCode>Sample-Routine</ActivityTypeCode>
    <ActivityMediaName>Water</ActivityMediaName>
    <ActivityStartDate>2006-02-02</ActivityStartDate>
    + <ActivityStartTime>
    <ActivityEndDate>2006-02-02</ActivityEndDate>
    + <ActivityEndTime>
    <ActivityRelativeDepthName>Midwater</ActivityRelativeDepthName>
    + <ActivityDepthHeightMeasure>
    <ActivityDepthAltitudeReferencePointText>SURFACE</ActivityDepthAltitudeReferencePointText>
    <ProjectIdentifier>RIBEACH-001</ProjectIdentifier>
    <MonitoringLocationIdentifier>RIBEACH-001-1</MonitoringLocationIdentifier>
  </ActivityDescription>

```

Figure 6. <ActivityDescription> data elements.

The <MonitoringLocationIdentifier> data element is only conditionally required because there are a few activity types that do not require a monitoring location, such as certain types of blanks (which are quality control samples), which can be tied to specific activities using Activity groups. The *Water Quality Exchange Domain Values List* names all activity types that can be used in the <ActivityTypeCode> data element and notes whether each one requires a monitoring location to be specified.

The <ActivityTypeCode> data element is required. As with monitoring locations, the *Water Quality Exchange Domain Values List* can be used to determine if an analytical method must be provided, based on the activity type code specified. For example, a routine sample requires the entry of an analytical method, while a field measurement does not.

In addition to the <ActivityTypeCode>, you must provide information in the following data elements: <ActivityIdentifier>, <ActivityMediaName>, <ActivityStartDate>, and <ProjectIdentifier>. Please note that the activity identifier must be unique across the organization, which is a change from STORET. To accommodate this stricter requirement, WQX allows the use of up to 35 characters to identify the activity. The *Water Quality Exchange XML Data Dictionary* documents additional business rules concerning activities.

By expanding the <ActivityStartTime>, <ActivityEndTime>, and <ActivityDepthHeightMeasure> blocks, you can view additional time and depth information, as shown in Figure 7. For example, you can see when an activity started and ended and at what depth an activity occurred. Time-related information, such as <TimeZoneCode> (e.g., EST [Eastern Standard Time]) must be an allowable value from the TIME_ZONE table in the *Water Quality Exchange Domain Value Lists* (see Appendix A).

The same rule applies to the <MeasureUnitCode> data element—the value for measure unit code must be an allowable value from the MEASUREMENT_UNIT table in the *Water Quality Exchange Domain Value Lists*. Please note that if the <ActivityDepthHeightMeasure> block is used, the <ActivityTopDepthHeightMeasure> block may not be used, and vice versa. This is because <ActivityDepthHeightMeasure> allows the user to specify a single measurement of depth or height, whereas the <ActivityTopDepthHeightMeasure> indicates that a measurement was taken over a range of depths. The *Water Quality Exchange XML Data Dictionary* provides additional business rules concerning activity measures.

```
- <ActivityStartTime>
  <Time>10:45:00</Time>
  <TimeZoneCode>EST</TimeZoneCode>
</ActivityStartTime>
<ActivityEndDate>2006-02-02</ActivityEndDate>
- <ActivityEndTime>
  <Time>11:00:00</Time>
  <TimeZoneCode>EST</TimeZoneCode>
</ActivityEndTime>
<ActivityRelativeDepthName>Midwater</ActivityRelativeDepthName>
- <ActivityDepthHeightMeasure>
  <MeasureValue>3</MeasureValue>
  <MeasureUnitCode>ft</MeasureUnitCode>
</ActivityDepthHeightMeasure>
```

Figure 7. More detailed <ActivityDescription> data elements.

Sample Description

The <SampleDescription> block is required for any activity that is some type of sample. Figure 8 shows an expanded view of the <SampleDescription> block. There are two requirements for the sample description, if it is included. First, the <SampleCollectionMethod> block must be provided. Second, the <SampleCollectionEquipmentName> must contain a value from the *WQX Domain Values List*. Please refer to the *Water Quality Exchange Data Dictionary* for other business rules for the <SampleDescription> block.

```

- <Activity>
+ <ActivityDescription>
- <SampleDescription>
  - <SampleCollectionMethod>
    <MethodIdentifier>ROUTINE</MethodIdentifier>
    <MethodIdentifierContext>21RIBCH_WQX</MethodIdentifierContext>
    <MethodName>Routine Sampling</MethodName>
    <MethodDescriptionText>500 mL sterile Nalgene bottle for beach water collection.</MethodDescriptionText>
  </SampleCollectionMethod>
  <SampleCollectionEquipmentName>Water Bottle</SampleCollectionEquipmentName>
  - <SamplePreparation>
    - <SamplePreparationMethod>
      <MethodIdentifier>BCHSAMPLE</MethodIdentifier>
      <MethodIdentifierContext>21RIBCH_WQX</MethodIdentifierContext>
      <MethodName>Beach Water Sample</MethodName>
    </SamplePreparationMethod>
    <SampleContainerTypeName>Nalgene Bottle</SampleContainerTypeName>
    <SampleContainerColorName>Opaque</SampleContainerColorName>
    <ThermalPreservativeUsedName>Wet Ice (4 deg C)</ThermalPreservativeUsedName>
    <SampleTransportStorageDescription>Samples stored in iced cooler while in transit.</SampleTransportStorageDescription>
  </SamplePreparation>
</SampleDescription>
+ <Result>

```

Figure 8. <SampleDescription> data elements.

Result

Although <Result> would seem to be a major block, it is a part of the <Activity> primary component, as shown in Figure 9. Because <Result> is not a primary component, the entire <Activity> must be submitted each time that a result is inserted or updated into the data warehouse. By expanding one of the two <Result> blocks under <Activity>, you can view the details of that result.

```

- <Activity>
+ <ActivityDescription>
+ <SampleDescription>
- <Result>
  + <ResultDescription>
  + <ResultAnalyticalMethod>
  + <ResultLabInformation>
  </Result>
+ <Result>
</Activity>
+ <Activity>
+ <ActivityGroup>

```

Figure 9. <Result> data elements.

Result Description

The <ResultDescription> block, which is shown in Figure 10, is where results information is stored. Most results fields require that allowable values be selected from the *Water Quality Exchange Domain Value Lists* (see Appendix A), except for <ResultMeasureValue> and <PrecisionValue> (not shown). For certain characteristics, the <ResultMeasureValue> must also be filled from the *Water Quality Exchange Domain Value Lists*. For example, the characteristic “Flow, stream stage (choice list)” must have one of the following results: ABOVE NORMAL, DRY, FLOOD, INTERSTITIAL, LOW, NO FLOW, or NORMAL.

It should also be noted that <ResultMeasureValue> (Figure 10) is a conditionally required field paired with <ResultDetectionConditionText> (Figure 11). For most applications in the Beaches user community, there are two cases: samples that produce a numerical value, and those that do not. For samples that produce a numerical value in <ResultMeasureValue>, <ResultDetectionConditionText> must be blank. For samples that do not produce a numerical value, <ResultMeasureValue> must be blank, and <ResultDetectionConditionText> must contain a value, such as “Not Detected” or “Present Below Quantification Limit”. Text values may also be used in the <ResultMeasureValue> field with similar business rules, however Beaches users usually report numerical result values. If <ResultDetectionConditionText> contains text, then the <DetectionQuantitationLimitMeasure> block (See Figure 13) must contain the detection limit or quantitation limit of the analytical method or instrument used by the laboratory. Figure 11 shows the <ResultDescription> for a case where <ResultDetectionConditionText> contains a value.

```

- <ResultDescription>
  <CharacteristicName>Fecal Coliform</CharacteristicName>
  - <ResultMeasure>
    <ResultMeasureValue>13</ResultMeasureValue>
    <MeasureUnitCode>MPN/100ml</MeasureUnitCode>
  </ResultMeasure>
  <ResultStatusIdentifier>Final</ResultStatusIdentifier>
  <StatisticalBaseCode>Mean</StatisticalBaseCode>
  <ResultValueTypeName>Calculated</ResultValueTypeName>
</ResultDescription>

```

Figure 10. <ResultDescription> data elements.

```

- <ResultDescription>
  <ResultDetectionConditionText>Not Detected</ResultDetectionConditionText>
  <CharacteristicName>Fecal Coliform</CharacteristicName>
  - <ResultMeasure>
    <MeasureUnitCode>MPN/100ml</MeasureUnitCode>
  </ResultMeasure>
  <ResultStatusIdentifier>Final</ResultStatusIdentifier>
  <StatisticalBaseCode>Mean</StatisticalBaseCode>
  <ResultValueTypeName>Calculated</ResultValueTypeName>
</ResultDescription>

```

Figure 11. <ResultDescription> data elements for non-detect.

The *Water Quality Exchange Domain Value List* for Characteristics also provides guidance on which characteristics require the <ResultSampleFractionText> field be filled in. Chemical characteristics usually require that a sample fraction be entered, but not always. Non-chemical characteristics usually do not require that a sample fraction be entered.

Result Analytical Method

The <ResultAnalyticalMethod> reusable data block is required only for certain activity types. In general, all types of samples (except depletion replicate samples) require an analytical method to be entered from the *Water Quality Exchange Domain Value Lists*. Field measurements never require an analytical method. In the example file that was provided with this training guide, allowable values from the *Water Quality Exchange Domain Value Lists* are used for the data elements shown in Figure 12, which include <MethodIdentifier>, <MethodIdentifierContext>,

and <MethodName>. However, if you input a non-domain value in <MethodIdentifierContext>, such as your state agency abbreviation, then you can define your own <MethodIdentifier> and <MethodName>. Please refer to the *Water Quality Exchange Data Dictionary* for complete business rules concerning the result analytical method.

```

+ <ResultDescription>
- <ResultAnalyticalMethod>
  <MethodIdentifier>9221-E</MethodIdentifier>
  <MethodIdentifierContext>APHA</MethodIdentifierContext>
  <MethodName>Fecal Coliform Procedure- Multiple-Tube Procedure</MethodName>
</ResultAnalyticalMethod>
+ <ResultLabInformation>

```

Figure 12. <ResultAnalyticalMethod> data elements.

Result Lab Information

The <ResultLabInformation> block, which is shown in Figure 13, contains information about the laboratory that conducted the analysis.

```

- <ResultLabInformation>
  <LaboratoryName>XYZ Laboratories</LaboratoryName>
  <AnalysisStartDate>2006-02-03</AnalysisStartDate>
- <AnalysisStartTime>
  <Time>09:00:00</Time>
  <TimeZoneCode>EST</TimeZoneCode>
</AnalysisStartTime>
- <ResultDetectionQuantitationLimit>
  <DetectionQuantitationLimitTypeName>Method Detection Level (MDL)
  </DetectionQuantitationLimitTypeName>
- <DetectionQuantitationLimitMeasure>
  <MeasureValue>2</MeasureValue>
  <MeasureUnitCode>MPN/100ml</MeasureUnitCode>
</DetectionQuantitationLimitMeasure>
</ResultDetectionQuantitationLimit>
</ResultLabInformation>

```

Figure 13. <ResultLabInformation> data elements.

The data values <TimeZoneCode> and <MeasureUnitCode> must be allowable values from the *Water Quality Exchange Domain Value Lists*. The <ResultLabInformation> block also contains a reusable data block called <DetectionQuantitationLimitMeasure>, or the detection limit of the analytical method used by the laboratory. This detection limit or quantitation limit must be provided if a <ResultDetectionConditionText> of “Not Detected,” “Present Above Quantification Limit,” or “Present Below Quantification Limit” is entered. The <DetectionQuantitationLimitTypeName> must always be provided along with detection or quantitation limit values. This field defines the type of detection or quantitation limit used in the analysis. In this example, it is “Method Detection Level (MDL)” because the analyte was below the level that could be detected by the method used.

Trip Blank Activity

Sometimes, you may want to provide information about QC samples, such as trip blanks. The primary purpose of blanks is to trace sources of artificially introduced contamination, such as

those due to shipping or laboratory procedures. The <Result> blocks of a trip blank will contain the results of this trip blank analysis.

As shown in Figure 14, a trip blank activity will look similar to other activities; however the <ActivityTypeCode> tag must say “Quality Control Sample–Trip Blank” and, as always, the <ActivityIdentifier> must contain a unique ID. Please note that the <SampleDescription> block is included for a trip blank activity because it is a type of sample and must contain a <SampleCollectionMethod>, but the <MonitoringLocationIdentifier> is not required. Results of the trip blank analysis are entered into each of the <Result> blocks. A final requirement for trip blanks is that they must be associated with the other related samples by creating an activity group. Examples of activity groups can be found at http://www.epa.gov/storet/wqx_products/Activity_Group_Examples.pdf.

```

- <Activity>
  - <ActivityDescription>
    <ActivityIdentifier>01-23-456-78-TB</ActivityIdentifier>
    <ActivityTypeCode>Quality Control Sample–Trip Blank</ActivityTypeCode>
    <ActivityMediaName>Water</ActivityMediaName>
    <ActivityStartDate>2006-02-02</ActivityStartDate>
  - <ActivityStartTime>
    <Time>10:45:00</Time>
    <TimeZoneCode>EST</TimeZoneCode>
  </ActivityStartTime>
  <ActivityEndDate>2006-02-02</ActivityEndDate>
  - <ActivityEndTime>
    <Time>11:00:00</Time>
    <TimeZoneCode>EST</TimeZoneCode>
  </ActivityEndTime>
  <ActivityRelativeDepthName>Midwater</ActivityRelativeDepthName>
  - <ActivityDepthHeightMeasure>
    <MeasureValue>3</MeasureValue>
    <MeasureUnitCode>ft</MeasureUnitCode>
  </ActivityDepthHeightMeasure>
  <ActivityDepthAltitudeReferencePointText>SURFACE</ActivityDepthAltitudeReferencePointText>
  <ProjectIdentifier>RIBeach-001</ProjectIdentifier>
</ActivityDescription>
  - <SampleDescription>
    + <SampleCollectionMethod>
      <SampleCollectionEquipmentName>Water Bottle</SampleCollectionEquipmentName>
    + <SamplePreparation>
  </SampleDescription>
  + <Result>
</Activity>

```

Figure 14. Example trip blank activity.

Activity Group

Activity groups are categories of activities, and include field set, replicate, subsample, and QC sample. For instance, to group a trip blank or other QC sample with other activities, you must create an activity group. Within an <ActivityGroup> block, activities are grouped by a unique activity group identifier, and all of the included activities (e.g., <ActivityIdentifiers> 01-23-456-78 and 01-23-456-78-TB) share a common <ActivityGroupTypeCode>, which must be an allowable value from the *Water Quality Exchange Domain Value Lists*. These fields are shown in Figure 15 for the case of a trip blank.

```
- <ActivityGroup>
  <ActivityGroupIdentifier>CBC-001-QC</ActivityGroupIdentifier>
  <ActivityGroupTypeCode>QC Sample</ActivityGroupTypeCode>
  <ActivityIdentifier>01-23-456-78</ActivityIdentifier>
  <ActivityIdentifier>01-23-456-78-TB</ActivityIdentifier>
</ActivityGroup>
```

Figure 15. <ActivityGroup> data elements.

Updating and Inserting Data into WQX

The schema that WQX uses to perform inserts and updates of data is shared between both of these operations because they are performed in a similar fashion. To perform either of these operations, the header document must contain one header and one payload. The payload must contain the tag <Payload Operation="Update-Insert"> for both of these operations and must reference WQX_WQX_v1.0.xsd as the schema. The WQX system will automatically determine whether a BEACH monitoring data submission is being updated or inserted.

Insert Process

In the case of inserting BEACH monitoring data into WQX, the data being inserted are new, and therefore not already present in WQX. All insert files must contain the complete root component for the entire hierarchy: Organization. Only one organization may be contained in the submission. Additionally, a BEACH monitoring data submission to WQX must contain at least one of the four following primary components:

- <Project>
- <Monitoring Location>
- <Activity>
- <Activity Group>

These primary components must be uniquely identified within an organization; however, results do not need to be uniquely identified within an organization. These are only unique within a parent activity. Within an activity, you may insert as many BEACH monitoring results as necessary.

Update Process

In the case of updating BEACH monitoring data in WQX, the data being updated are already present in WQX. As with the insert process, an update submission file must contain only one complete organization, which is the root component, and at least one of the four following primary components:

- <Project>
- <Monitoring Location>
- <Activity>
- <Activity Group>

To be updated, these primary components must already be uniquely identified within an organization. The primary component level is where all BEACH monitoring data must be updated. This is because children of the primary components are not uniquely identified within

the organization. For instance, if you want to update Result 1 of Activity 2, you must update all of Activity 2, because Activity 1 may also contain a Result 1.

Certain general rules apply to updates of BEACH monitoring data that are already in WQX. Following these rules will help you avoid unintended changes to your data during updates. Table 3 provides a list of these rules.

Data that contain errors will cause the load to fail at the primary component level. This means that if a result fails because of an error, the entire parent activity will fail to load, along with all associated child results. Other activities with no errors will load fine because all successful loads are immediately committed into the database. The following are two approaches that you may use to correct failed loads to WQX:

- Correct the data that failed and resubmit the entire job. Those records that successfully loaded will be automatically treated as an “Update” and the failed record will be treated automatically as an “Insert.”
- Correct the data that failed and submit only the failed portions. The advantage of this approach is that your file may be significantly smaller; therefore, it will process more efficiently.

Table 3. General Rules that Apply to Updates of BEACH Monitoring Data.

Intended Update Action	Rule
Update only the primary component, but not child data. Example: Update activity, but leave associated results alone.	Provide complete primary component tags (e.g., activity tags), but do not include results tag.
Update only the primary component and delete child data. Example: Update activity and delete all associated results.	Provide complete primary component tags (e.g., activity tags) and include an empty Results tag.
Update all results in an activity.	Provide complete activity tags and complete tags with updated information for all results in that activity.
Update only one of multiple results in an activity.	Provide complete activity tags and complete tags with information for all results in that activity with updated information for the one result to be updated.
Update one of multiple results in an activity and delete all other results.	Provide complete activity tags and updated information for only the result to be updated. Do not provide information for results to be deleted.

Deleting Data from WQX

The schema that WQX uses to perform deletes of data allows the deletion of data for an entire primary component. In order to perform a “Delete,” the header document must contain one header and one payload. The payload must contain the tag <Payload Operation=“Delete”>, and must reference WQX_WQX_Delete_v1.0.xsd as the schema.

Delete Process

If you discover that BEACH monitoring data are no longer valid, it is necessary to remove that data from WQX by performing a delete. The WQX Delete schema allows you to extract data

from WQX at the primary component level. For example, if you want to delete an entire project, monitoring location, activity, or activity group, you would use the Delete schema. However, if you want to remove one result from within an activity, you should follow the “Update-Insert” process instead.

The payload section of the header document for a deletion from WQX must contain the <OrganizationIdentifier> and one or more of the following: <ProjectIdentifier>, <MonitoringLocationIdentifier>, <ActivityIdentifier>, and <ActivityGroupIdentifier>. The <OrganizationIdentifier> is included to provide context for the delete. Only one organization can be provided. Organizations cannot be deleted from WQX using the Delete schema. The primary components can be repeated, thus allowing the deletion of several projects, monitoring locations, activities, or activity identifiers at one time. Figure 16 shows an example of a complete WQX XML file used to delete BEACH monitoring data for the project RIBEACH-001 and the monitoring location RIBEACH-001-1 from the organization 21RIBCH_WQX.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <Document Id="UI34" xmlns="http://www.exchangenetwork.net/schema/v1.0/ExchangeNetworkDocument.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
- <Header>
  <Author>John Doe</Author>
  <Organization>RIDOH</Organization>
  <Title>WQX</Title>
  <CreationTime>2007-01-03T12:43:18</CreationTime>
  <ContactInfo>John Doe 401-555-1234</ContactInfo>
</Header>
- <Payload Operation="Delete">
- <WQXDelete xmlns="http://www.exchangenetwork.net/schema/wqx/1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.exchangenetwork.net/schema/wqx/1
http://www.exchangenetwork.net/schema/wqx/1/0/WQX_WQX_Delete_v1.0.xsd">
  <OrganizationIdentifier>21RIBCH_WQX</OrganizationIdentifier>
  <ProjectIdentifier>RIBEACH-001</ProjectIdentifier>
  <MonitoringLocationIdentifier>RIBEACH-001-1</MonitoringLocationIdentifier>
</WQXDelete>
</Payload>
</Document>
```

Figure 16. Delete XML file.

Please note that whenever a parent primary component is deleted, all child data are also deleted. The delete XML file, shown in Figure 16, will also remove the activities that are associated with the project and monitoring location being deleted. Other examples of this include the following:

- Attached Binary Objects are deleted when a parent project, monitoring location, or activity are deleted.
- Results are deleted when the parent activity is deleted.
- Activities are deleted when their parent project or monitoring location is deleted. The exceptions to this are when an activity is related to more than one project or monitoring location, and also, as in the case of QC samples, when it does not relate to any monitoring location.

Summary

The purpose of this tutorial is to help you understand the data submission format and XML tags and blocks that are required to create a WQX XML file to submit BEACH monitoring data to the STORET warehouse. In addition, the tutorial illustrates how to handle trip blanks and activity groups. The information contained in the tutorial is based on the *Water Quality Exchange XML*

Flow Configuration Document, the *Water Quality Exchange Data Dictionary*, and the *Water Quality Exchange Domain Values Lists*, which are available with additional information at http://www.epa.gov/STORET/future_storet.html, and on the National Environmental Information Exchange Network's Web site at <http://www.exchangenetwork.net>.

Based on this introduction to WQX and XML, you can determine the best way to generate an XML file for your BEACH monitoring data. Most users will use Oracle, Visual Basic, or other software programs to generate XML files directly from their data system. For advice on this process or for more information, please call the STORET hotline at 1-800-424-9067, or you can e-mail support staff at storet@epa.gov.

[This page intentionally left blank.]

Appendix A

The following table shows XML tags and the allowable values that are associated with them. For example, for the <ActivityMedia>, you are required to select air, water, biological tissue, soil, sediment, or other. For initial data mapping purposes, a Microsoft Access database of WQX domain values can be downloaded from http://www.epa.gov/storet/future_downloads.html. The table below provides the domain table and domain field names that must be used to fill each data element XML tag. Once WQX is running, the “WQX.GetDomainValueByElementName” Web service method will also allow you to query or solicit current domain value lists. These services will return an XML file in the format of the schema WQX_DomainValues_v1.0.xsd referred to in the *Water Quality Exchange Flow Configuration Document*. An example XML file that would be returned by the Web service can be found at http://www.exchangenetwork.net/schema/WQX/1/WQX_DomainValues_XMLExample_v1.0.xml.

Data Element XML Tag	WQX.GetDomainValueByElementName Service Parameter	Domain Table	Domain Field
ActivityGroupTypeCode	ActivityGroupType	ACTIVITY_GROUP_TYPE	AGTYP_NAME
ActivityMediaName	ActivityMedia	ACTIVITY_MEDIA	ACMED_NAME
ActivityMediaSubDivisionName	ActivityMediaSubDivision	ACTIVITY_MEDIA_SUBDIVISION	AMSUB_NAME
ActivityRelativeDepthName	ActivityRelativeDepth	RELATIVE_DEPTH	RELDPTH_NAME
ActivityTypeCode	ActivityType	ACTIVITY_TYPE	ACTYP_CD
AddressTypeName	AddressType	ADDRESS_TYPE	ADDTYP_NAME
CharacteristicName	Characteristic	CHARACTERISTIC	CHR_NAME
CountryCode	Country	COUNTRY	CNTRY_CD
CountyCode	County	COUNTY	CNTY_FIPS_CD
DetectionQuantitationLimitTypeName	DetectionQuantitationLimitType	DETECTION_QUANT_LIMIT_TYPE	DQLTYP_NAME
ElectronicAddressTypeName	ElectronicAddressType	ELECTRONIC_ADDRESS_TYPE	EATYP_NAME
HorizontalCollectionMethodName	HorizontalCollectionMethod	HORIZONTAL_COLLECTION_METHOD	HCMTH_NAME
HorizontalCoordinateReferenceSystem DatumName	HorizontalCoordinateReferenceSystem Datum	HORIZONTAL_REFERENCE_DATUM	HRDAT_NAME
MeasureQualifierCode		RESULT_MEASURE_QUALIFIER	RMQLF_CD
MeasureUnitCode	MeasureUnit	MEASUREMENT_UNIT	MSUNT_CD
MethodIdentifier		ANALYTICAL_METHOD	ANLMTH_ID
MonitoringLocationTypeName	MonitoringLocationType	MONITORING_LOCATION_TYPE	MLTYP_NAME
ResultDetectionConditionText	ResultDetectionCondition	RESULT_DETECTION_CONDITION	RDCND_NAME
ResultLaboratoryCommentCode	ResultLaboratoryComment	RESULT_LAB_COMMENT	RLCOM_CD
ResultMeasureValue (must match a domain list value if the CharacteristicName ends with the phrase “Choice List”)	ResultMeasureValuePickList	CHARACTERISTIC_PICK_LIST_VALUE	CPLVAL_CD
ResultSampleFractionText	ResultSampleFraction	SAMPLE_FRACTION	SMFRC_NAME
ResultStatusIdentifier	ResultStatus	RESULT_STATUS	RESSTA_NAME
ResultTemperatureBasisText	ResultTemperatureBasis	RESULT_TEMPERATURE_BASIS	RTMPB_NAME
ResultTimeBasisText	ResultTimeBasis	RESULT_TIME_BASIS	RTIMB_NAME
ResultValueTypeName	ResultValueType	RESULT_VALUE_TYPE	RVTYP_NAME
ResultWeightBasisText	ResultWeightBasis	RESULT_WEIGHT_BASIS	RWBAS_NAME
SampleCollectionEquipmentName	SampleCollectionEquipment	SAMPLE_COLLECTION_EQUIP	SCEQP_NAME

Data Element XML Tag	WQX.GetDomainValueByElementName Service Parameter	Domain Table	Domain Field
SampleContainerColorName	SampleContainerColor	CONTAINER_COLOR	CONCOL_NAME
SampleContainerTypeName	SampleContainerType	CONTAINER_TYPE	CONTYP_NAME
SampleTissueAnatomyName	SampleTissueAnatomy	SAMPLE_TISSUE_ANATOMY	STANT_NAME
SampleTissueTaxonomicName	Taxon	TAXON	TAX_NAME
StateCode	State	STATE	ST_CD
StatisticalBaseCode	StatisticalBase	RESULT_STATISTICAL_BASE	RSBAS_CD
TelephoneNumberTypeName	TelephoneNumberType	PHONE_TYPE	PHTYP_NAME
ThermalPreservativeUsedName	ThermalPreservativeUsed	THERMAL_PRESERVATIVE	THPRSV_NAME
TimeZoneCode	TimeZone	TIME_ZONE	TMZONE_CD
TribalCode	Tribe	TRIBE	TRB_CD
VerticalCollectionMethodName	VerticalCollectionMethod	VERTICAL_COLLECTION_METHOD	VCMTH_NAME
VerticalCoordinateReferenceSystem DatumName	VerticalCoordinateReferenceSystem Datum	VERTICAL_REFERENCE_DATUM	VRDAT_NAME