CHAPTER 1

Getting Started with the XML Engine

What Does the XML Engine Do?

The XML engine processes an *XML document*. The engine can:

- export (write to an output location) an XML document from a SAS data set of type DATA by translating the SAS proprietary file format to XML markup. The output XML document can then be:
  - used by a product that processes XML documents.
  - moved to another host for the XML engine to process by translating the XML markup back to a SAS data set.
- import (read from an input location) an external XML document. The input XML document is translated to a SAS data set.

Understanding How the XML Engine Works

Assigning a Libref

The XML engine works much like other SAS engines. That is, you execute a LIBNAME statement in order to assign a libref and specify an engine. You use that libref throughout the SAS session where a libref is valid.

A libref for the XML engine can be associated with either a specific XML document or, for the XML92 engine nickname, the physical location of a SAS library in a directory-based environment. You can use a wildcard character in document names, such as **`*.xml`**. When you use the libref, SAS either translates the data in a SAS data set into XML markup, or translates the XML markup into SAS format.

Importing an XML Document

To import an XML document as a SAS data set, the following LIBNAME statement assigns a libref to a specific XML document and specifies the XML engine:

```
libname myxml xml 'C:\My Files\XML\Students.xml';
```

Executing the DATASETS procedure shows that SAS interprets the XML document as a SAS data set:

```
proc datasets library=myxml;
```

Output 1.1 PROC DATASETS Output for MYXML Library

```
                                             Directory

           Libref         MYXML
           Engine         XML
           Physical Name  C:\My Files\XML\Students.xml
           XMLType        GENERIC
           XMLMap         NO XMLMAP IN EFFECT


                                                         Member
                                      #  Name      Type

                                      1  STUDENTS  DATA
```

The PRINT procedure results in the following output:

```
proc print data=myxml.students;
run;
```

Output 1.2 PROC PRINT Output of SAS Data Set MYXML.STUDENTS

```
                                The SAS System

     Obs    STATE     CITY           ADDRESS            NAME                  ID

       1    Texas     Huntsville     1611 Glengreen     Brad Martin         1755
       2    Texas     Houston        11900 Glenda       Zac Harvell         1522
```

Exporting an XML Document

To export an XML document from a SAS data set, the LIBNAME statement for the XML engine assigns a libref to the XML document to be created.

In the following code, the first LIBNAME statement assigns the libref MYFILES to the SAS library that contains the SAS data set Singers. The second LIBNAME statement assigns the libref MYXML to the physical location of the XML document that is to be exported from Myfiles.Singers:

```
libname myfiles 'C:\My Files\';

libname myxml xml 'C:\My Files\XML\Singers.xml';
```

Executing these statements creates the XML document named Singers.XML:

```
data myxml.Singers;
   set myfiles.Singers;
run;
```

Output 1.3 XML Document Singers.XML

```
<?xml version="1.0" encoding="windows-1252" ?>
<TABLE>
   <SINGERS>
      <FirstName> Tom </FirstName>
      <Age> 62 </Age>
   </SINGERS>
   <SINGERS>
      <FirstName> Willie </FirstName>
      <Age> 70 </Age>
   </SINGERS>
   <SINGERS>
      <FirstName> Randy </FirstName>
      <Age> 43 </Age>
   </SINGERS>
</TABLE>
```

SAS Processing Supported by the XML Engine

The XML engine supports the following processing:

- □ The XML engine supports input (read) and output (create) processing. The XML engine does not support update processing.
- □ The XML engine is a *sequential access* engine in that it processes data one record after the other. The engine starts at the beginning of the file and continues in sequence to the end of the file. The XML engine does not provide random (direct) access, which is required for some SAS applications and features. For example, with the XML engine, you cannot use the SORT procedure or ORDER BY in the SQL procedure. If you request processing that requires random access, a message in the SAS log notifies you that the processing is not valid for sequential access. If this message occurs, put the XML data into a temporary SAS data set before you continue.

Transferring an XML Document Across Environments

When you transfer an XML document across environments, for example using FTP, you must be aware of the document's content in order to determine the appropriate transfer mode. If the document contains either an encoding attribute in the XML declaration or if a byte order mark precedes the XML declaration, transfer the file in binary mode. If the document contains neither criteria and you are transferring the document across similar hosts, transfer the file in text mode.

When you export an XML document using the XML engine, by default, the XML document contains an encoding attribute in the XML declaration from the SAS data set's encoding, for example, **<?xml version="1.0" *encoding*="*windows-1252*" ?>**. You can override the SAS data set's encoding when you export the XML document by specifying the XMLENCODING= LIBNAME statement option.

Frequently Asked Questions

Is the XML Engine a DOM or SAX Application?

The XML engine uses a Simple API for XML (SAX) model, not a Document Object Model (DOM). SAX does not provide a random-access lookup to the document's contents; it scans the document sequentially and presents each item to the application one item at a time.

Does the XML Engine Validate an XML Document?

The XML engine does not validate an input XML document. The engine assumes that the data passed to it is in valid, well-formed XML markup. Because the engine does not use a DTD (Document Type Definition) or SCHEMA, there is nothing to validate against.

What Is the Difference between Using the XML Engine and the ODS MARKUP Destination?

Typically, you use the XML engine to transport data, while the ODS MARKUP destination is used to create XML from SAS output. The XML engine creates and reads XML documents; ODS MARKUP creates but does not read XML documents.

Why Do I Get Errors When Importing XML Documents Not Created with SAS?

The XML engine reads only files that conform to the markup types supported in the XMLTYPE= LIBNAME statement option. Attempting to import free-form XML documents that do not conform to the specifications required by the supported markup types will generate errors. To successfully import files that do not conform to the XMLTYPE= markup types, you can create a separate XML document, called an XMLMap. The XMLMap syntax tells the XML engine how to interpret the XML markup into a SAS data set or data sets, variables (columns), and observations (rows).

See Chapter 3, "Importing XML Documents," on page 25, Chapter 5, "Importing XML Documents Using an XMLMap," on page 47, "LIBNAME Statement Syntax" on page 93, and Chapter 9, "Creating an XMLMap," on page 105.