

XML Basics

What is it?

Extensible Markup Language (XML) is an open standard language used to create files for exchanging and displaying data. XML is an outgrowth of Standard Generalized Markup Language and provides a standard method for describing data based upon a syntax developed by the World Wide Web Consortium (W3C). XML enables the definition, transmission, and validation of data between applications and between organizations. It is the primary data format supported for data exchange by the Environmental Information Exchange Network (Exchange Network).

XML uses markup symbols to describe and define data in a readable format. XML allows users to define metadata which, through the use of tags, identifies data elements contained within a document. This helps facilitate information exchange between systems by ensuring that all parties exchanging data share a common understanding of what is being exchanged. In addition to these new advances, XML can handle any operation of its predecessor, Hypertext Markup Language (HTML), which focuses primarily on how the text on the Web page looks (e.g., whether the text is bold, underlined, or justified to the right). XML is very useful for organizations that do not share a common set of data standards, but need to understand the data that they are sharing.

How does it work?

Development of XML documents begins with the identification and definition of the data elements that will be displayed or exchanged. The data elements are defined using tags that indicate what a data element represents. For example, in the simplest explanation a person's name might contain three tags:

```
<Name>  
  <First>Fred</First>  
  <Middle_Initial>M</Middle_Initial>  
  <Last>Smith</Last>  
</Name>
```

If an XML document contains a large number of data elements, or tags, then developers may create schema to define the names and rules for the tags contained in the application.

XML schema are the files that serve as the framework for defining the data elements and rules in an XML document. Schema express shared vocabularies and allow computers to carry out rules made by people. They provide a means for defining the structure, content and semantics of XML documents. These data structures are designed by the exchange partners and reviewed for consistency with data standards and best practices by the State/EPA Technical Resource Group. Schema can be attached to an XML document or

may exist as a separate document. In addition to creating custom schema, a variety of schema are readily available on the Web for developers to integrate into their own documents.

In EPA, schema are being developed for each major environmental program collection. Within these collections are reusable components that can be shared across programs. The Exchange Network then uses these schema as a common language to exchange data using Network Nodes on the Exchange Network. These schema can be found on the www.exchnagenetwork.net web site.

How does this impact the Environmental National System processing?

Instead of Regional, State and local agencies mapping their databases to the current column delimited format, they translate to the defined XML format based on the schema. These agencies would have the option to submit in XML format by either developing these translators for their own systems or recoding their databases to receive and process XML in its original format.

Schema are drafted for many of the EPA national systems modules. After completion of EPA and State review, testing, and revision, agencies will have the option to submit data in XML via a web-based upload process. Following this XML deployment will be the use of nodes to conduct machine-to-machine exchanges across the Exchange Network. No determination has been made to eliminate the existing formats. Rather, the XML option will allow States that are ready to submit XML to have this capability.

How does XML benefit the EPA National System Programs?

XML allows programs at EPA and other Federal, State, and local agencies to share and analyze data sets that cross program boundaries. By building in data standards into the industry accepted XML format, all programs use the same predetermined tag names and conventions for same data elements. Some stakeholders may have a need or desire to analyze data that cross program boundaries such as the Air Quality System Program, RCRAInfo or Toxics Release Inventory Program. By standardizing data types and formats across programs, duplicate data collection can be reduced. In addition, XML can establish additional rules for submitting data. This can increase the quality of the data transaction and reduce the manual correction process on the receiving end.

How can I find out more about XML or the Exchange Network?

World Wide Web Consortium Web Site

<http://www.w3.org>

Exchange Network Web Site

<http://www.exchangenetwork.net>

Environmental Council of States

<http://www.sso.org/ecos>

EPA Central Data Exchange Public Web Site

<http://www.epa.gov/cdx>

EPA Data Standards Web Site

<http://www.epa.gov/edr>

Network Grants EPA Regional Contacts

<http://www.epa.gov/neengprg/contacts.html>