



Exchange Network Document Validation Services

Introduction

The XML schema validator is a set of XML web services for validating XML documents against the associated schemas and custom rules. It may consist of two steps:

1. Schema Validator: This service verifies the structure of XML documents using definitions in one or more schema files. Basic content constraints are also checked. **Step One is mandatory and must be completed prior to submitting via CDX-Web site or the CDX Node. If you are submitting XML files, you are required to complete a successful validation of your file(s) using the EPA authorized validation process. Your file(s) will not be considered accepted by EPA if this validation process has not been completed.**
2. Schematron Validator: This is a prototype service that validates XML documents to accomplish additional quality control checks on the data using custom rules, lookup tables, and regular expressions. During the NEI 2004 calendar year submission period, while this additional service remains in prototype, and possibly limited availability, you are not required to complete a successful validation. However if you do encounter this service as also available when you are performing step 1 above, it means that the schematron is considered fully operable. Using it will help you identify some basic content errors that you can correct prior to completing your submission to EPA. These additional quality checks will benefit you and EPA by reducing the time and costs required to otherwise identify and correct such errors in a post-submission process.

For more information contact the CDX Help Desk at 888-890-1995 and Press Two for Network Node or email NodeHelpDesk@csc.com.

File Formats

The service accepts two file formats: XML or ZIP. A compressed document is always preferred, and it is mandatory when a document exceeds the 2 MB set threshold. The ZIP file must contain only one XML document. For security reasons, the service may reject the submission of a very large uncompressed document. In this case, there is no return response

Using the Validator

You can send an XML document for verification from your desktop using the following procedure:

1. Open your web browser and go to <http://tools.epacdxnode.net>
2. Enter `http://tools.epacdxnode.net/xml/validator.wsdl` in the WSDL File Address box, and click the Retrieve button. This brings up HTML forms as shown below:

SOAP Web Client - Microsoft Internet Explorer

Address: http://naas.epacdxnode.net/soapclient?CacheWSDL=0&template=%2Fclientform.html&fn=soapform&SoapTemplate=%2F5oapResult.html&S

Network Node Client Forms

Service Documentation: A set of xml validation services for the National Environmental Information Exchange Network (NEIEN)

SOAP Method : SchemaValidate

Documentation: Validate an XML document using XML schemas. sendResultTo should be your email address if the XML document is large. It is also a good idea to zip the XML Document

Server Address: http://naas.epacdxnode.net/xml/validator.wsdl

documentType: NEI_AreaNonroad_v3_0

xmlDocument: Browse...

docFormat: xml

sendResultTo: xsd:string

Show: Response **Format:** XML **Invoke**

SOAP Method : SchematronValidate

Documentation: Validate an XML document using schematron rules. This operation should be performed only if your document passes the XML schema validation.

Server Address: http://naas.epacdxnode.net/xml/validator.wsdl

documentType: NEI_AreaNonroad_v3_0

xmlDocument: Browse...

docFormat: xml

sendResultTo: xsd:string

Show: Response **Format:** XML **Invoke**

3. Always select **Soap Method: SchemaValidate** first (the first method on the web page).
4. Click on the **Browse** button to choose the document to be validated on your computer. It can either be an XML document, or a ZIP file that contains the document.
5. Select the **documentFormat (docFormat)** from the dropdown list. It should be either XML or ZIP.
6. Enter your email address in the **sendResultTo** box. Multiple email addresses, separated by ';', are also accepted. The service will send validation results to you through email. If your document is too large to be completed in a short period of time, a background job will be started. This is so called Asynchronous Processing Mode (see description below).
7. Select **HTML** as the output Format.
8. Click the **Invoke** button. This will send the document for validation. You will receive a response message.

The following figure shows an example for validating a zipped NEI Point submission:

SOAP Method : SchemaValidate

Documentation: Validate an XML document using XML schemas. `sendResultTo` should be your email address if the XML document is large. It is also a good idea to zip the XML Document

Server Address: `http://naas.epacdxnode.net/xml/validator.wsdl`

documentType:

xmlDocument:

docFormat:

sendResultTo: `xsd:string`

Show: **Format:**

Synchronous and Asynchronous Processing

When a document is larger than a pre-configured threshold, the service will notify user that the request has been received, and it will be processed in an asynchronous mode. A valid email address (the `sendResultTo` parameter) is required; otherwise you will not be able to receive the results.

Common Error Conditions

1. Invalid XML Document: When a non-XML document is received but marked as XML format, the server returns:
Error at Line 1, position 1 Error c00cee2d, Incorrect document syntax.
Or,
Error at Line 1, position 0 Error c00cee3a, A document must contain exactly one root element.
This also happens when sending a ZIP file but the `docFormat` parameter is XML.
2. Invalid ZIP File: When a ZIP file is received but the service is unable to unzip it, it returns a SOAP format message with the following description:
The file is either corrupted or is not in a valid zip format.
3. Invalid Document Type: When an incorrect schema (the `documentType` parameter) is chosen. The service returns the following error:
Validate failed because the root element had no associated DTD/schema.

This usually indicates there is a mismatch between the schema and the XML instance document.

4. Invalid Numerical Data: An element of numerical type, such as integer and double, cannot have an empty value (i.e., with no value or NULL). The element should be removed if optional. The common error returned under such error may be something like:

Error parsing '' as double datatype.

Note from Testing the Schematron Validator for NEI Documents

Numeric (number or decimal) fields in the NEI input format (NIF) may not validate properly against the complimentary validator if the tags are sent as empty tags. This is due to a difference in the development validator (XMLSpy) which is also a common local validator in general, and the validation used at the Schematron level. The exception are the start and end time elements which have been pattern matched in a different manner than the other numeric fields.

There are three approaches to manage the null data/empty non-mandatory tag issue:

1. If the submitting agency never collects the information, do not include that tag in the mapping.
2. If the submitting agency sometimes (but not always) collects the information, include logic in the local XML instance document generation programming to prevent the output of empty tags."
3. Depending upon the XML instance documentation generation method that the submitting agency is using, the solution in #2 may be prohibitive. In that case, it is recommended that the empty tags be removed from the document using a find/replace edit method. The XML instance document will need to be opened (should be copied first to a backup) with an editor (XMLSpy or MS Word (using Text form) are all methods that may work even on large files) and the find/replace mode will need to be used. This method may leave an empty line in the document, however this has been tested against the complimentary validator, and does not cause a problem.

It is hoped that the number of fields which may actually required management through approach #2 or #3 will be a manageable subset of the following worksheet.