|  |
| --- |
| Data Interchange Document XML Schema Validation |
| Validation of XML data interchange documents with normative schemas using free, open-source software (XML Explorer) |
| A product of |

2012-16-02

Contents

[Section 1: Introduction 1](#_Toc319045840)

[Section 1.1: What is schema validation? 1](#_Toc319045841)

[Section 1.2: Data exchange services 2](#_Toc319045842)

[Section 1.2.1: What is a web service? 2](#_Toc319045843)

[Section 1.2.2: What is USGIN? 2](#_Toc319045844)

[Section 1.2.3: Where can I find NGDS schemas? 3](#_Toc319045845)

[Section 2: Schema Validation 3](#_Toc319045846)

[Section 2.1: What you need: 3](#_Toc319045847)

[Section 2.2: Accessing the appropriate schema 3](#_Toc319045848)

[Section 2.3: Performing a GetFeature Request 3](#_Toc319045849)

[Section 2.3.1: GetFeature requests and XML Documents 3](#_Toc319045850)

[Section 2.3.2: A sample GetFeature Request 4](#_Toc319045851)

[Section 2.3.3: Breaking down a GetFeature request 5](#_Toc319045852)

[Section 2.4: Downloading the results of your GetFeature request 5](#_Toc319045853)

[Section 2.5: XML Schema Validation with XML Explorer 6](#_Toc319045854)

[Section 3.0: Troubleshooting 7](#_Toc319045855)

# Section 1: Introduction

One of the objectives of the US Geoscience Information Network (USGIN) is to establish a community of practice for developing, documenting, adopting, and using standard document encoding for data interchange. The technology for encoding schemes is evolving continuously, and current practice is to use XML encoding with data schema defined by XML schema.

Validation is a software process to verify that a given XML document conforms to some XML schema document. It is easier to write reliable software utilizing XML interchange formats if the schema for the XML documents that are being used is known, and the instance documents being processed validate against that schema. This document provides basic information regarding the protocols and software requirements for schema validation.

This document was written primarily intended for Association of American State Geologists (AASG) subrecipients subcontracted by the Arizona Geological Survey (AZGS) under the Department of Energy (DOE) contract Number DE-EE0002850 for the National Geothermal Data System (NGDS) to perform the subcontract requirement to make "at risk" geothermal data available online to promote geothermal development throughout the United States. This data is being made available using USGIN practices and is a stepping stone to building the network.

## Section 1.1: What is schema validation?

An XML schema is a kind of XML document that uses XML syntax to define the structure of other XML documents, which will be referred to as XML instance documents. The basic schema definition consists of a collection of elements (analogous to entities, objects or tables in other data modeling paradigms), each of which has a set of properties. Each property has an associated data type (which might be another XML element), and cardinality that specifies how many instances of a property may be associated with the containing element. Schema validation is the process of determining if the elements and properties in an XML instance document conform to the definitions in a schema document. Schema validation provides a layer of quality control and assurance for XML data interchange documents. Conformance to a schema greatly facilitates writing reliable software that will utilize information contained in the interchange documents.

## Section 1.2: Data exchange services

This document is not primarily intended to describe schema validation at an abstract level. Rather, it is written with an eye for validating data that is accessed in XML interchange documents using community XML schema. Specific examples in this tutorial use data accessed from DOE National Geothermal Data System (NGDS) web feature services (WFS). The Association of American State Geologists Geothermal Data Project is tasked with compiling and publishing data on the WWW. Following USGIN practice, common, high value datasets have been identified for distribution using community XML interchange formats, to allow easy integration of data provided from various state and federal data servers in the system. The interchange formats are specified by XML schema, which are developed by first specifying the content model as a table in an MS Excel spreadsheet (see [http://www.stategeothermaldata.org/­data\_delivery/­content\_model\_templates](http://www.stategeothermaldata.org/data_delivery/content_model_templates) ), and then implementing the content model as an XML schema (see <http://schemas.usgin.org/schemas/> ). When NGDS partners publish these data, it is expected that the data exchange documents obtained from **web features services** will validate against the XML schema adopted for that particular data type.

### Section 1.2.1: What is a web service?

A web service is a web-accessible resource that offers some collection of operations invoked using a documented protocol (see the [USGIN glossary entry describing web services](http://usgin.org/content/glossary" \l "web_service)). This protocol specifies syntax for requesting each operation, including definition of what input parameters are required, what changes will be effected on the server by the operation, and what kind of output will be returned to the client that invokes the operation. An important part of this protocol is the message format used to encode information sent to the server in a request, and to encode information returned by the server. In our example cases here, the message format of particular interest is the XML encoding used to transmit requested data back to the client computer. By using standardized requests and responses, client software can make requests for data regardless of server configuration. The service defines an interface a client uses to interact with a server; this interface decouples the client and server, allowing the two systems to develop independently without having to re-engineer their interaction.

This tutorial focuses on XML data interchange documents that are returned by Open Geospatial Consortium web feature services (WFS), which are currently used by USGIN and the State Geothermal Data project for data exchange.

### Section 1.2.2: What is USGIN?

The **United States Geoscience Information Network (USGIN,** <http://usgin.org/>) is a partnership between the Association of American State Geologists (AASG) and the United States Geological Survey (USGS) created to facilitate discovery of, and access to, geoscience information provided by state and federal geological surveys of the United States. USGIN is a distributed data-sharing network defined by a collection of open-source standards and protocols for sharing geoscience data, and a community that uses these conventions for data registration, publication, discovery, and use. USGIN is implemented using existing World Wide Web infrastructure, and promotes the use of free, open-source software where applicable. Anyone can access the data shared over USGIN, and any data that is shared according to USGIN specifications is automatically a part of the network.

### Section 1.2.3: Where can I find NGDS schemas?

NGDS XML schemas can be found at: <http://schemas.usgin.org/schemas/>

# Section 2: Schema Validation

## Section 2.1: What you need:

1. A computer with an Internet connection
2. A **web service** that will return XML instance documents to validate
3. XML **schema** documents to which the instance is intended to conform. Schema for XML documents are usually stored using the .xsd file extension. As mentioned in Section 1.2.1, AASG schemas are located at http://schemas.usgin.org/schemas/.
4. A software package that does XML schema validation. This document provides instructions for schema validation using XML Explorer (available at: <http://xmlexplorer.codeplex.com/>)
5. Basic understanding of XML syntax, namespaces and data types.

## Section 2.2: Accessing the appropriate schema

Your first step is to inspect the root element in an xml document obtained from the service you are testing. Since we are using OGC WFS as the basis for this tutorial, this root element will be wfs:FeatureCollection. The root element will have a collection of attributes that define namespaces for xml elementsThe next step in the schema validation process will be to download the schema against which you will validate your web service.

1. In your web browser navigate to <http://schemas.usgin.org/schemas/>
2. Click the appropriate schema
   1. The appropriate schema will usually correspond to the type of features in your web service
      1. For example: if your web service contains WellHeaders, then you will use the WellHeader schema to validate the web service
   2. If you are unsure of the appropriate schema for your web service, contact the Arizona Geological Survey
3. After the schema appears in your web browser, click the **File** menu in your web browser and click **Save Page As…**
   1. Note that some modern web browsers hide the menu bar by default; you might need to press the Alt key on your keyboard to temporarily reveal the menu bar
4. Save the schema to a directory on your computer you intend to use for schema validation
   1. XML schema documents use the .xsd file extension

## Section 2.3: Performing a GetFeature Request

To validate your web service against the schema you just downloaded, you will need to submit a **GetFeature** request to your desired web service. **To perform a GetFeature request, simply enter your GetFeature request into a web browser in the same manner as you would a standard URL.**

A getFeature request is a **WFS request** – that is, a request for data from a web feature service (WFS). A GetFeature request returns an XML representation of the attributes of features in the web service.

Note: large web services can have tens of thousands of features, so an unfiltered GetFeature request can be very demanding in terms of bandwidth and system resources. When entering GetFeature requests into a web browser, it is possible to filter the results of the request by appending conditions to the GetFeature request, but doing so is difficult because the filter syntax for GetFeature requests is complicated.

### Section 2.3.1: GetFeature requests and XML Documents

As indicated in Section 2.2, a GetFeature request will return an **XML representation of the attributes of features in the web service**. This indicates the following:

* A GetFeature request returns an **XML document**
* The XML document provided in response to a GetFeature request is a representation of **features**
* A feature in a web service is a cartographic representation of a real-world object
* Features are described by **attributes**
* Attributes include data such as latitude and longitude coordinates and any other information relevant to a feature

For example: a web feature service might contain fifty features representing river systems in the United States. The attributes describing each feature might include the latitude and longitude coordinates of each river system, as well as information such as flow rates, seasonal volume, and depth at the river’s deepest point.

An XML representation of the above example would list each feature in the web service, as well as associated attributes, within the structure of an XML document.

A GetFeature request submitted to the web feature service in the above example would return the XML document described above.

For more information about XML, visit the following locations on the USGIN website:

* [XML](http://usgin.org/content/glossary#xml)
* [Markup language](http://usgin.org/content/glossary#markup_language)
* [Element](http://usgin.org/content/glossary#element)
* [XML Tutorial](http://usgin.org/content/xml-tutorial/)

### Section 2.3.2: A sample GetFeature Request

The following is a sample GetFeature request:

**http://services.azgs.az.gov/arcgis/services/aasggeothermal/CAWellHeaders/MapServer/WFSServer?service=WFS&request=GetFeature&typeName=Wellheader**

If entered into a web browser, this request will return an XML document representing all **Wellheader** features in the web service (Figure 1).

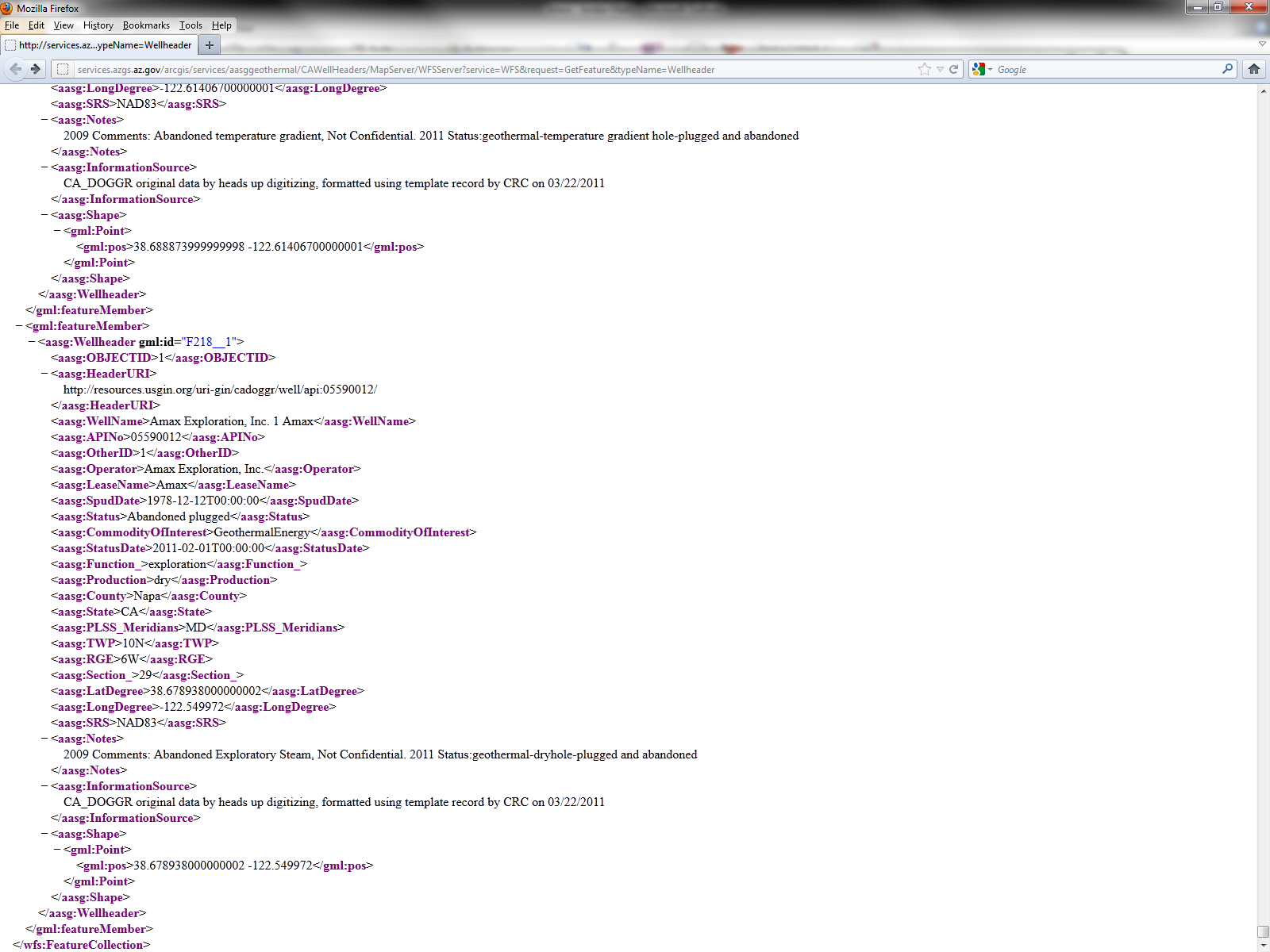


Figure 1

### Section 2.3.3: Breaking down a GetFeature request

A GetFeature request can be broken down into two component parts: the **service endpoint**, and the **request** proper. These are demonstrated in the example below:

**http://services.azgs.az.gov/arcgis/services/aasggeothermal/CAWellHeaders/MapServer/WFSServer?service=WFS&request=GetFeature&typeName=Wellheader**

In this example, the **red** text constitutes the **service endpoint**; the **green** text constitutes the **request**.

The **service endpoint** is the web location that handles the service request (http://services.azgs.az.gov/arcgis/services**)**. This is followed by the directory and folder containing the Web Feature Service ( WFS) (/aasggeothermal/CAWellHeaders**)** The service endpoint can be further subdivided into individual [tokens](http://usgin.org/content/glossary#token), each of which can vary according to web service. A full explanation of the tokens in the service endpoint is beyond the scope of this document.

The **request** includes the desired **operation** and a collection of parameters that control the operation; some parameters are optional and others are required. In the above example, the feature requested is “Wellheader.”

## Section 2.4: Downloading the results of your GetFeature request

Having made a GetFeature request in your web browser, your next step is to download the XML response (hereafter referred to as the **GetFeature document**) to your computer and prepare to validate it. To prepare your GetFeature document for validation, you will need to modify it slightly after you download it.

1. Having performed a GetFeature request in your web browser, navigate to the **File** menu in your web browser and click **Save Page As…**
   1. This will save the GetFeature document to your computer as an XML document
   2. XML documents use the .xml file extension
   3. Note that some modern web browsers hide the menu bar by default; you might need to press the **Alt** key on your keyboard to reveal the menu bar
2. Save the **GetFeature document** **to the same directory on your computer to which you downloaded the schema in Section 2.2**
3. In Windows, navigate to the directory in which you just downloaded the **GetFeature document**
4. Open the **GetFeature document** you just downloaded in a text editor such as **Notepad** or **Wordpad**
   1. This can typically be accomplished by **right-clicking** the document and clicking **Open With…** in the context menu that appears
   2. Note that some text editors are easier to use for viewing XML documents than others – Notepad, for example, contains no features to make XML documents more human-readable. Consider trying different text editors to find the most optimal combination for viewing XML documents. Recommended text editors include:
      1. WordPad: basic text editor included with Windows
      2. Notepad++: an actively maintained free-and-open-source text editor designed with XML support
5. Near the very top of the **GetFeature document**, highlight the **service URL** (Figure 2)

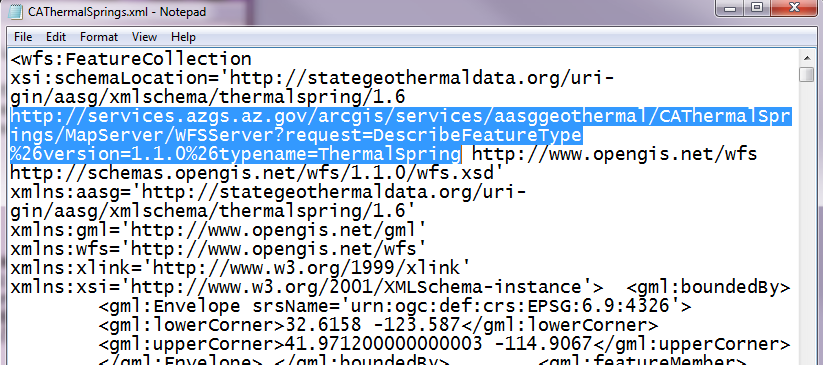


Figure 2

1. Replace the **service URL** with the exact name of the schema you just downloaded, complete with file extension (Figure 3).

## Section 2.5: XML Schema Validation with XML Explorer

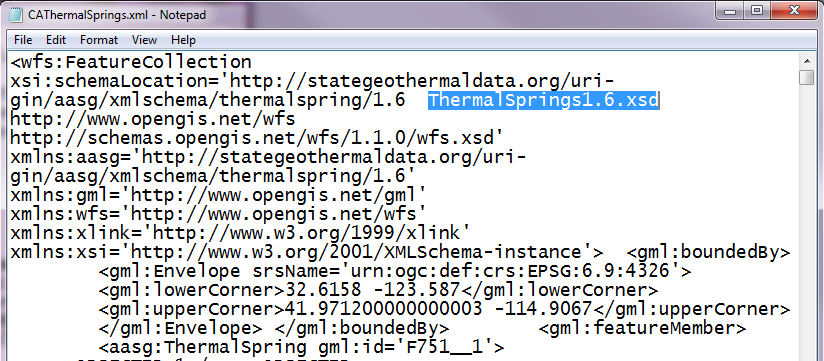


Figure 3

Having downloaded your schema document and your GetFeature document, and having prepared your GetFeature document for validation, open the GetFeature document in XML Explorer.

1. Open XML Explorer
2. Click the **File** menu and click **Open**
3. Navigate to and open your **GetFeature document**
4. Click the **Errors** tab (Figure 4)
5. If the web service validates successfully against the designated schema, no errors should appear and your web service is valid

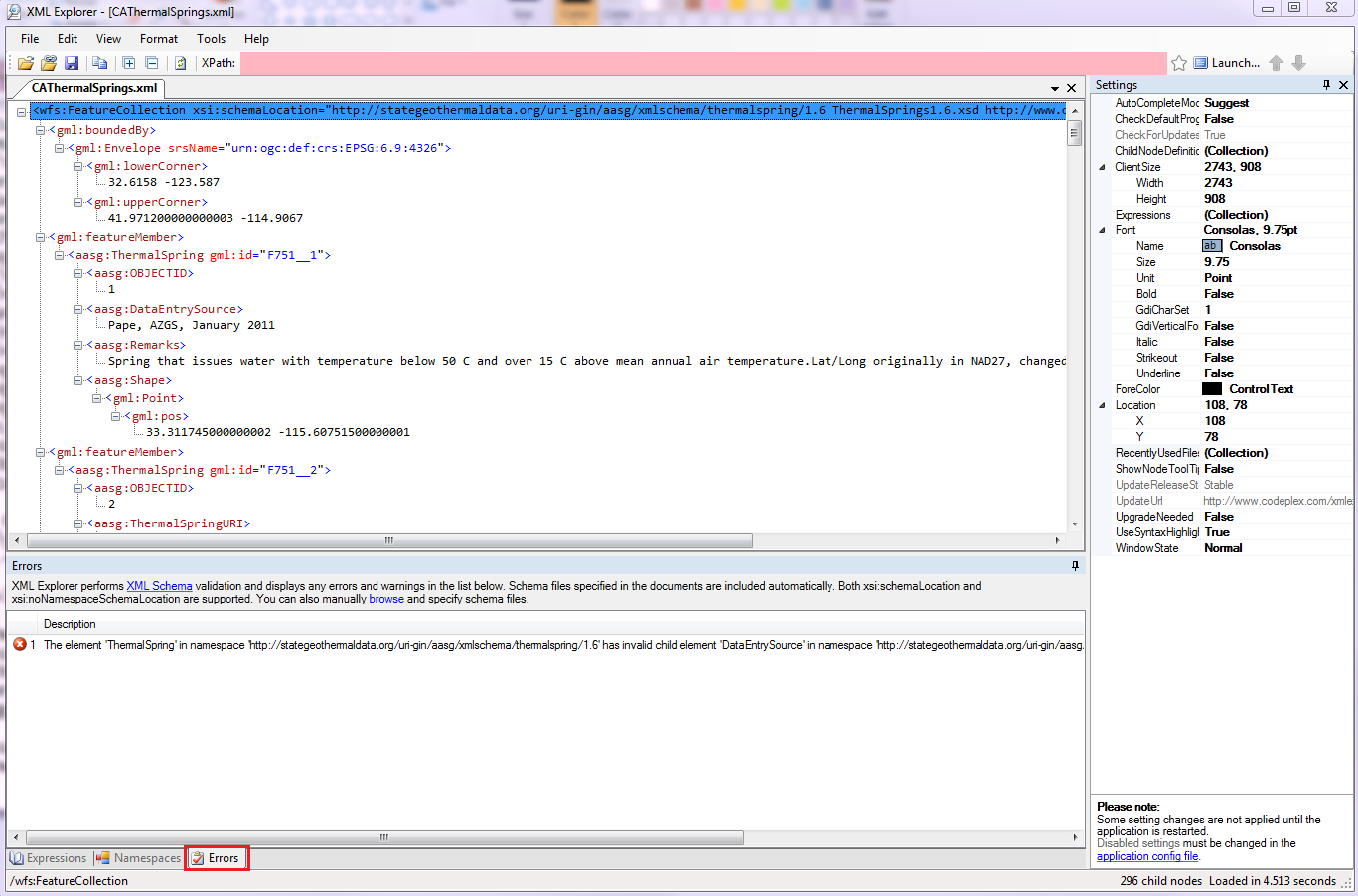


Figure 4

If errors appear, you will need to fix them within the web service itself. The process for doing so is outside the scope of this document.

# Section 3.0: Troubleshooting

For questions on this process, the schemas, or other options for xml validation, please feel free to contact Celia Coleman at [celia.coleman@azgs.az.gov](mailto:celia.coleman@azgs.az.gov).