Simple Feature Content Model Guidelines

August 29, 2013 SMR, CMC

Move from NGDS GDSDPWG Sharepoint to USGIN Specs directory. These guidelines are being proposed for use in USGIN for broader model development and documentation.

This is a ‘living document’ that will be revised and updated as practice evolves.

# Introduction

This document is intended to provide guidance for creating an Excel workbook that defines a content model for a flat file data interchange format. The content model defines the information that will be associated with a feature or observation type; the content model may be implemented in a variety of ways, but USGIN is currently implementing these interchange formats as GML Simple Features to be served by an OGC WFS.

For point data, the Excel worksheet can be converted to a GIS feature class that can be used to deploy the WFS. The worksheet can thus serve as a means to package and check datasets for distribution through a service.

If the data require line or poly feature classes, then the actual data interchange template is NOT in the spreadsheet, but in a geodatabase feature class because the geometry or shape field can not be implemented (in a useful way…) in the spreadsheet. Recommended practice is that the Excel workbook should be named xxxContentElementsN.N.xls, where xxx is the feature name. This worksheet should contain a field list that defines the thematic (not spatial geometry) fields in the content model. The data compilation/delivery package for line or polygon features (Fault, contact, geologicUnitOutcrop in GeologicMapData, activeFault contains the content elements workbook, a personal geodatabase containing a single feature class (or multiple feature classes representing each layer that will appear in the final service) that is the template for deploying the service, or a zip archive with shape file and excel workbook with one sheet for the template and a readme saying this is a last resort if you can’t work with geodatabase. – see ActiveFault ( https://github.com/usgin-models/ActiveFaults) as an example.

Worksheet/Tab names in the workbook should be consistent and are discussed in the sections below. To inspect templates currently in use by the AASG geothermal data NGDS project, visit http://schemas.usgin.org/models/ . These will be posted following approval by the Geothermal Data System Development and Population Working Group.

# Worksheets

The two sheets that are expected to be completed by the data provider will be listed first; Data tab is followed by Metadata tab.

## Data

All templates should have xxxURI, xxxName, Notes (not remarks, or description—only one free text additional information field unless there’s a compelling reason for more…), and Source. ‘xxx’ is a prefix that ideally is the same as the feature name in the service to be deployed, and the same prefix is used for this worksheet name and the field containing the URI for the feature in the content model. xxxURI, xxxName and Source are always mandatory. Note second place version number for the content model will need to increment if field names are changed or fields added.

## Metadata

This worksheet includes information about the data provider that will be used to create the metadata record describing the dataset loaded into the spreadsheet. Required values are a title, description (abstract), and originator for the dataset, along with a telephone number or e-mail address for each one.

## FieldList

The FieldList worksheet provides a listing of all elements (fields) in the content model, specifies data types, xml implementation, and provides information to define the content and explain usage of the content element. The FieldList worksheet should have these columns:

### **Interchange Content Element**:

This is the field name that appears in column headings in the template worksheet, and will become the XML element name in the interchange format. Note that many of these element names are longer than 10 characters and will be truncated if the content model is implemented in a dBase table (e.g. as an ESRI Shapefile). The adherance to the naming conventions in the 7 fields in this tab is also important for developers. The text listed in this tab (Element Description, Domain, and Example(s)) should be also indicated in pop-up boxes for each field name in the Data tab.

### **Data Type**:

Logical data type. Valid Values:

* free text: any XML valid alpha numeric characters, no limitation on length of content.
* term: element value is a word from a controlled vocabulary. Implication is that there will be a list of terms in the DataValidTerms worksheet that may be used to populate this field. If users add vocabulary terms, they should be added to the DataValidTerms sheet and defined.
* date: a date and time value
* URI: a unique identifier as defined by IETF 3986 (<http://tools.ietf.org/html/rfc3986>)
* Decimal: a number that quantifies a value along a continuum.
* Integer: a number representing a discrete, countable quantity.

### **Implementation**:

XML data type used to implement the element.

* string: sequence of alphanumeric characters, no length restriction
* string ISO 8601: string conforming to ISO8601 syntax for date/time information. "yyyy-mm-ddThh:mm:ss"
* string nnn: sequence of alphanumeric characters with a maximum length of nnn characters
* double: a floating point number, used to represent real numbers
* integer: a number in the set {…-2, -1, 0, 1, 2,…}

### **Cardinality**:

Restriction on the number of instances of an element that may be present in a valid interchange document. For GML simple features, the valid values are ‘1’ (required) or ‘0..1’ (optional). To encode an attribute that has multiple values using a GML simple feature, the adopted approach is to concatenate the multiple values into a single string value with ‘|’ (pipe) character delimeters separating individual values. The individual values may be a tuple; syntax for these tuples is defined in the Element Description column for that attribute.

### **Element Description**:

Text description of content element. This text is required for each element. The text should define the intention of the element content for feature description, any limitations on the range of acceptable values, conventions to be followed in string syntax for content in the field. This text should mirror exactly the annotations for each field in the accompanying schema (XSD).

**Domain**

Text indicates appropriate domains for given fields, containing a number range or a list of terms for use in the template worksheet n some cases a complete, closed vocabulary may be supplied, in other cases the template may provide some sample values and data providers are expected to add any other terms they use, along with definitions of the terms.

**Example(s)**

**Provides examples of expected data content or formatting.**

## About

This is the title page for the workbook, and includes title of the content model, version number, description of the intention of the model, description of the feature or observation of interest, list of editors (contributors), and revision history information for the template (from NGDS developers).. Information about each of the tabs in the workbook also exist here.

DataRevisionsWorksheet meant to provide feedback for data providers to implement necessary edits to the data to conform to schema conventions, but is also meant to serve as a revision history for the data.

# Conventions

## Field names

Field names may not contain spaces. Use CamelCase if name has more than one word (first letter of each word capitalized, spaces removed. Underscores may be used if multiple capital letters appear in sequence. Special characters !@#$%^&\*(){}[]\|~`”’<>,/? may not be used in field names. Valid non-alphanumeric characters are ‘-‘, ‘\_’ and ‘.’ XML field names may not begin with a numeral.

If a field has dataType ‘URI’, the field name should have the suffix ‘URI’.

## Other conventions

The Excel outline feature may be used to hide groups of fields that have related content that is optional. Users are free to modify the outlining to hide cells that they are not using, but a column in the template spreadsheet should never be renamed or deleted.

Required field headings are highlighted in the Data tab and FeildList tab.

The field heading row is the first row of the Data tab. This row should be a frozen pane (Freeze Top Row under Review in Excel).