# State Geothermal Data Content Models Tutorial

<http://repository.usgin.org/>

The State Geothermal Data (SGD) project is an implementation of the Geoscience Information Network (GIN) model for a distributed, interoperable network of geologic data. SGD implements the rules, protocols and standards of GIN.

If you are a subrecipient under the Arizona Geological Survey (AZGS) for the Department of Energy (DOE) National Geothermal Data System (NGDS), and if you have data to submit for the SGD project, then you will be using the SGD content models to format your data for delivery. This tutorial will describe the usage of SGD content models.

Content models represent one method by which SGD facilitates interoperability. Content models provide a *schema*, or structural configuration, for submitted data. Schemas dictate where and how data should be entered. Content models use *templates*, preformatted spreadsheets, to make it easy for data providers to submit their data in accordance with interoperable schemas.

Note: To complicate matters, the terms *content model* and *template* are often used interchangeably in SGD and GIN documentation. Technically, the data delivery *template* is a preformatted spreadsheet that demonstrates a schema; the template spreadsheet is found within the content models. Content models are usually Excel workbooks containing multiple spreadsheets, one of which is the template spreadsheet. Unfortunately, sometimes these Excel workbooks (the content models) are referred to as template files, despite the fact that the template is only one spreadsheet in the workbook.

For more information about content models, schemas, and interoperability, go [here](#example).

TABLE OF CONTENTS

[State Geothermal Data Content Models Tutorial 1](#_Toc300062800)

[SUBMITTING DATA USING CONTENT MODELS 3](#_Toc300062801)

[Finding Content Models on the State Geothermal Data Website 5](#_Toc300062802)

[Finding Content Models in the USGIN Document Repository 7](#_Toc300062803)

[WORKING IN YOUR CONTENT MODEL 8](#_Toc300062804)

[ENTERING DATA 10](#_Toc300062805)

[SUBMITTING DATA 11](#_Toc300062806)

[SCHEMAS, CONTENT MODELS, TEMPLATES, AND INTEROPERABILITY: AN EXAMPLE 13](#_Toc300062807)

# SUBMITTING DATA USING CONTENT MODELS

As demonstrated in the [example](#example), content model templates are used to facilitate interoperability at the data level, making it easy for computers to use data from multiple sources and humans to analyze such data without interpretation.

Consequently, data submitted for the SGD project should be formatted in accordance with existing content models.

Workflow for the use of content models is described in Figure 1.

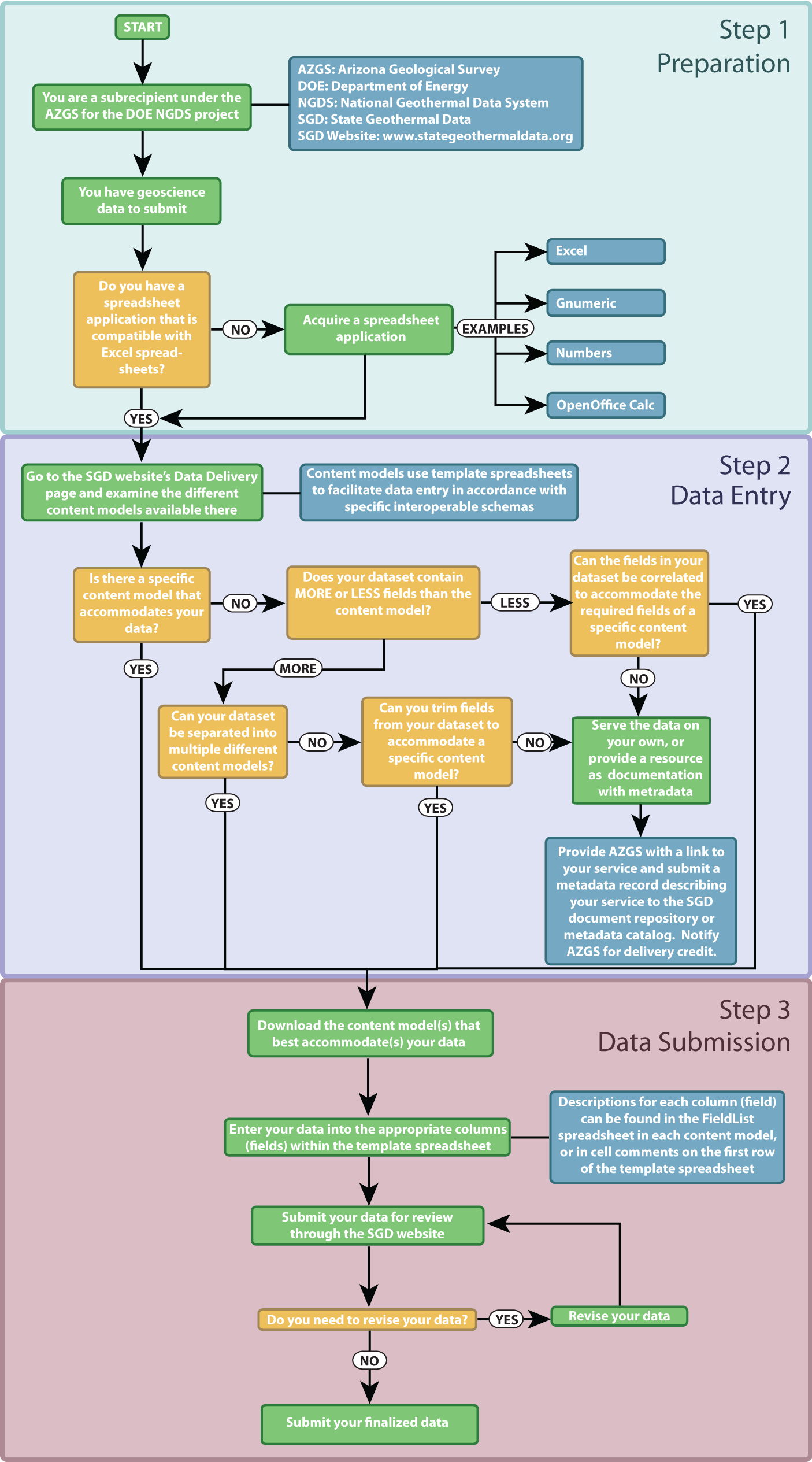


Figure : Content Model Workflow

In order for subrecipients to submit data for the SGD project via content models, subrecipients must first have data they wish to submit. In addition, this data must be accommodated by the content models. For example, SGD has no content models capable of supporting meteorological data. A list of content models can be found at the bottom of the page.

SGD content models are stored and submitted as Excel spreadsheets with the .xls file extension. Consequently, data providers must use a program capable of reading and writing Excel spreadsheets (Table 5).

|  |  |  |  |
| --- | --- | --- | --- |
| **Application** | **Developer** | **Compatible Operating Systems** | **Distribution** |
| Excel | Microsoft | Windows, Mac OSX | Proprietary |
| Gnumeric | GNOME Office Team | Windows, Mac OSX, Linux, BSD, Unix | Free and open-source |
| Numbers | Apple, Inc | Mac OSX | Proprietary |
| OpenOffice Calc | Oracle Corporation | Windows, Mac OSX, Linux, BSD, Unix | Free and open-source |

Table : Popular, actively maintained spreadsheet applications capable of reading and writing Excel spreadsheets

Assuming you are a subrecipient with data to submit; assuming your data can be accommodated by one of the content models listed at the bottom of the page; and assuming that you have an application capable of interacting with Excel spreadsheets, the next step is to locate and download an appropriate content model.

Content models can be located on both the State Geothermal Data website and the USGIN website. Both sites provide access to the same content models.

## Finding Content Models on the State Geothermal Data Website

1. Using your preferred browser, go to [www.stategeothermaldata.org](http://www.stategeothermaldata.org)
2. Mouse over the **Data Delivery** tab (Figure 2)
3. In the menu that appears, click **Content Model Templates** (Figure 2)
4. At the bottom of the Content Model Templates page are multiple tables containing content models and associated descriptions (Figure 3). These tables can searched by title or description, and they can be sorted by title or date on which they were last updated.
5. Click a link in the **Title** column of the table in Figure 3 to access the desired content model. This will take you to a corresponding USGIN Document Repository page containing metadata about your template and a download link for your template file (Figure 4).
6. Under Available Files (near the bottom of Figure 4), click the link to your template file; when prompted, choose the location on your computer to which you wish to download the template and save it there. It is recommended that you select the latest-greatest template version, associated with a higher version number (version 1.22 is more recent than version 1.0), as later versions are usually furnished with updated functionality.

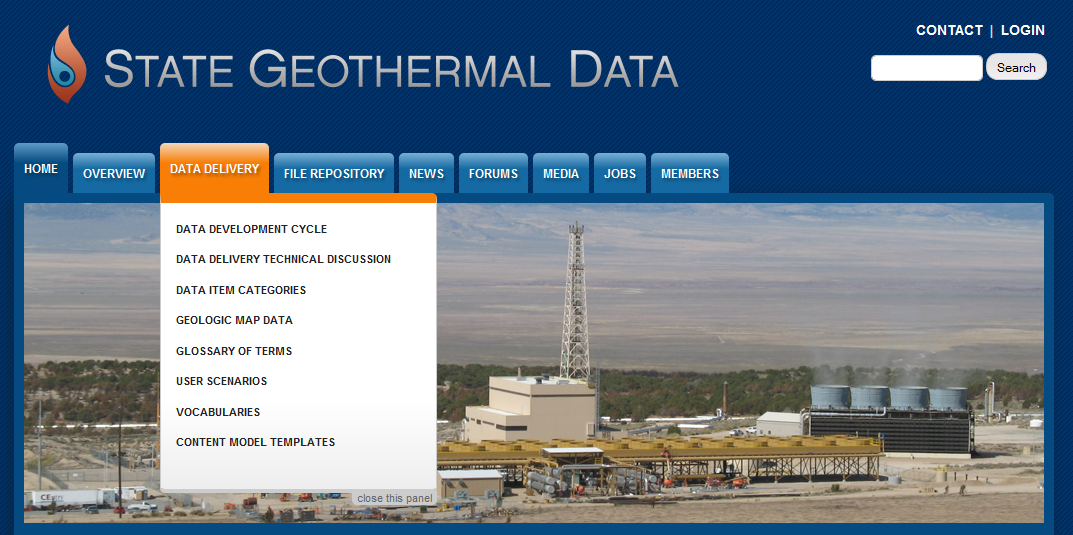


Figure : SDG Data Delivery

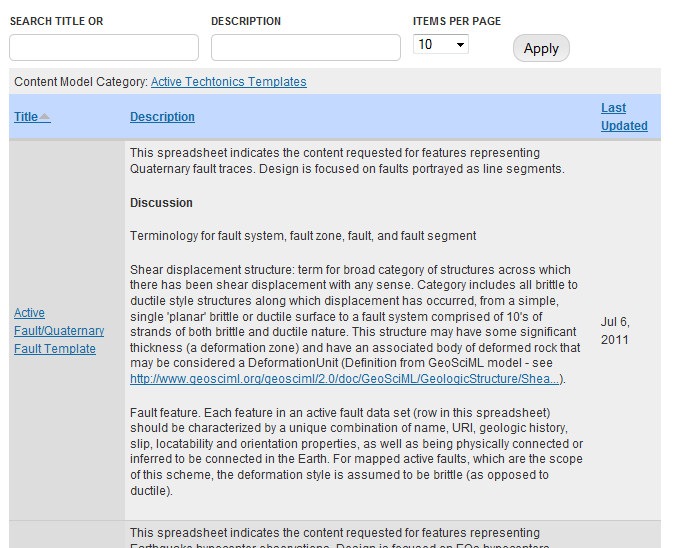


Figure : Templates on the Content Model Templates page

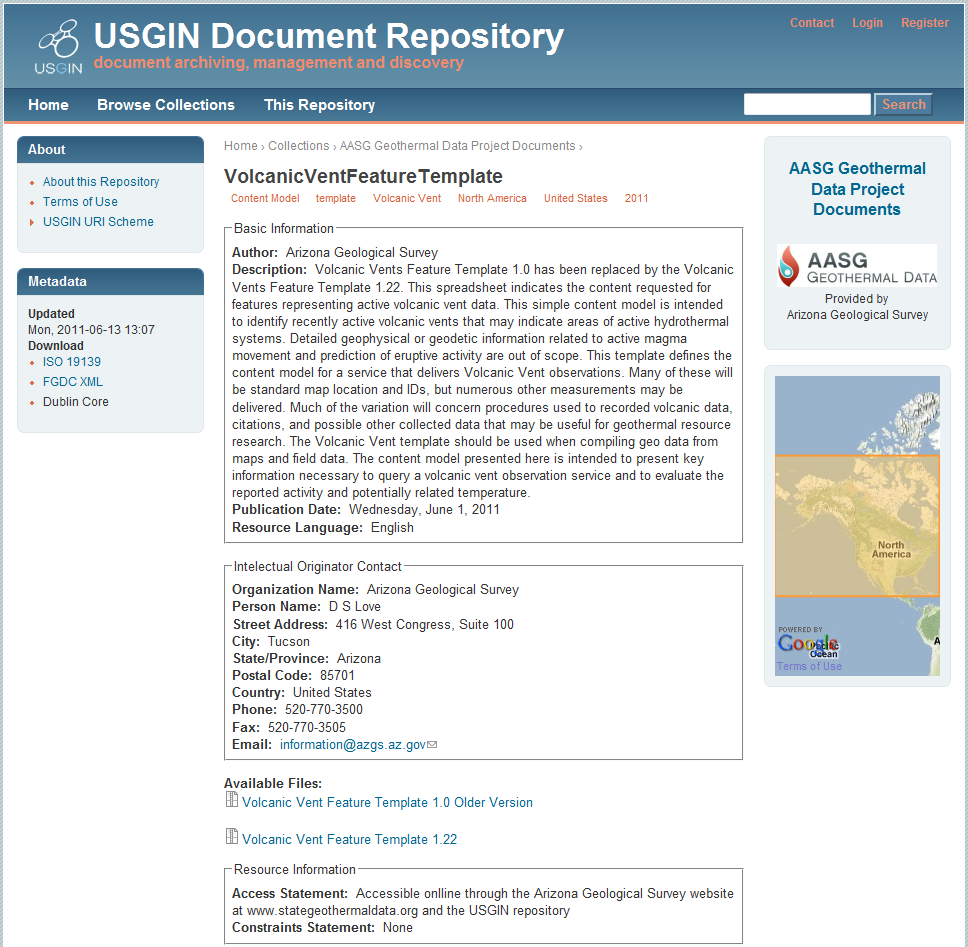
****

Figure : A USGIN Document Repository template page

## Finding Content Models in the USGIN Document Repository

1. Using your preferred browser, go to [repository.usgin.org](http://www.stategeothermaldata.com)
2. Click **Browse Collections** in the USGIN header (Figure 6)
3. On the Collections page, click **AASG Geothermal Data Project Documents** (Figure 7).
4. Browse the list and click on the appropriate template. This will take you to a corresponding USGIN Document Repository page containing metadata about your template and a link to your template file (Figure 5).
5. Under Available Files (near the bottom of Figure 5), click the link to your template file; when prompted, choose the location on your computer to which you wish to download the template and save it there. It is recommended that you select the latest-greatest template version, associated with a higher version number (version 1.22 is more recent than version 1.0), as later versions are usually furnished with updated functionality.

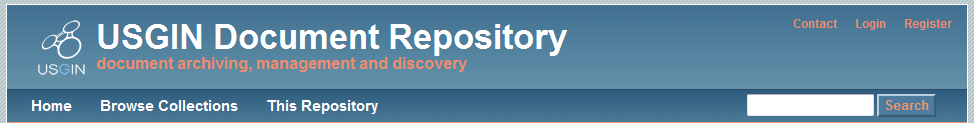


Figure : The USGIN Document Repository Header

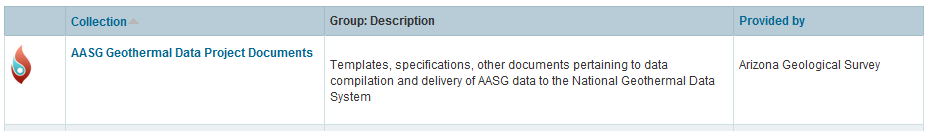


Figure : The SGD Collection

# WORKING IN YOUR CONTENT MODEL

Having found and downloaded a content model that meets your needs, find the content model on your computer and open it. The content model may be compressed as a ZIP file; if so, extract the content model from the ZIP file and open it.



Figure : Tabs in Excel; each tab indicates an individual spreadsheet

Each content model is an Excel workbook containing several spreadsheets; each spreadsheet is associated with its own tabs tabs (Figure 7)

* About: Describes the content model, its authors, version information and history.
* Notes: Information specific to a particular content model, including conventions, color codes, and specific terminology.
* Dataset Metadata: Information used to populate a metadata record for the completed dataset. In other words: once the content model is populated with actual data, the metadata for that data will be entered into the Dataset Metadata spreadsheet, from which an independent metadata record will eventually be generated. For more information about metadata, see the USGIN metadata tutorial and glossary entry.
* Template: The spreadsheet in which actual data will be entered (Figure 8) according to the practices defined in the Field List and Data Valid Terms spreadsheets.

The first row of information in the Template spreadsheet is used to generate *fields* in a database. Consequently, each column represents a *database field*, and columns in the Template spreadsheet are often referred to as fields.

Each column is intended to record only the data indicated in the cell at the top of the column. For example, if the cell at the top of column D is labeled APINo, then the column should contain only API numbers for each database record. For specific instructions regarding each field, see the Field List spreadsheet.

It should be noted that the content models often refer to columns or fields as elements or exchange elements. Don’t be alarmed! **The terms field and element are used interchangeably because they refer to different implementations of the same concept.** An element can be considered the XML equivalent of a database field or spreadsheet column.

* Field List: The Field List tab describes each field of the Template spreadsheet in detail (Figure 9), indicating the kind of data each field should contain and the manner in which that information should be entered. The Field List will also indicate whether or not specific vocabulary from the Data Valid Terms tab should be used in a given field.

It should be noted that not all Field List spreadsheets are at the same level of development – some will provide more information than others. Optimally, the Field List spreadsheet will provide a description for each field, as well as data entry instructions, notes explaining the peculiarities of a given field, and examples of data entered into the field.

* Data Valid Terms: Some fields, as described in the Field List spreadsheet, require very specific vocabulary; data entered in those fields should be restricted to that listed on the Data Valid Terms spreadsheet (Figure 11).
* Reviewer Comments: Comments from users regarding data entered in the content model.

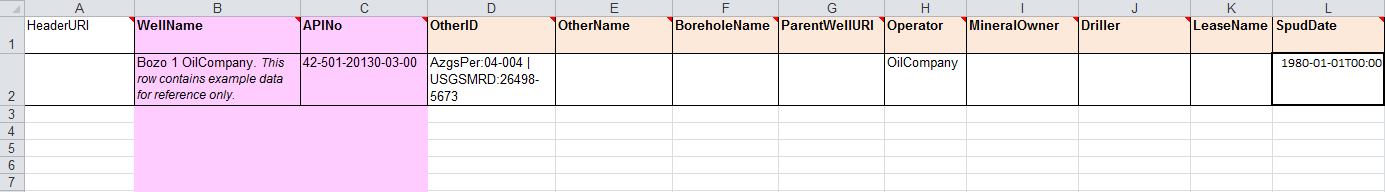


Figure : The Template spreadsheet. The cells at the top of each column contain information used to generate database fields.

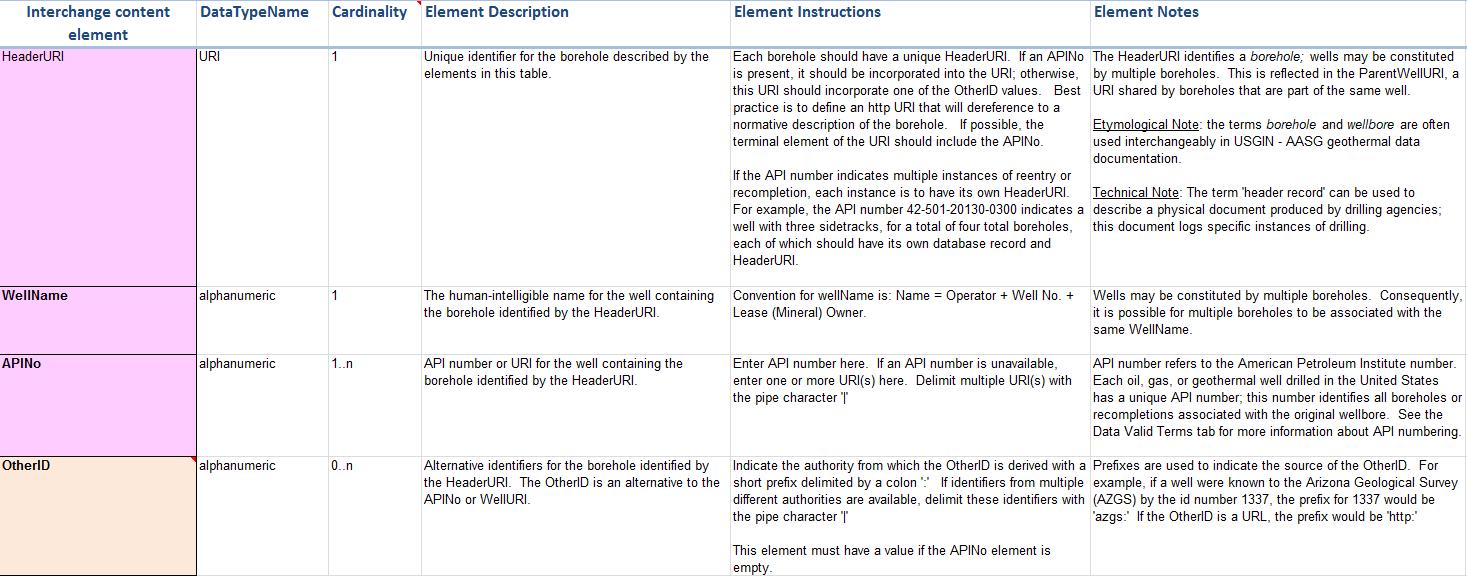


Figure : The Field List tab. The far left column corresponds with the topmost row in the Template spreadsheet; columns on the right provide descriptions for each field

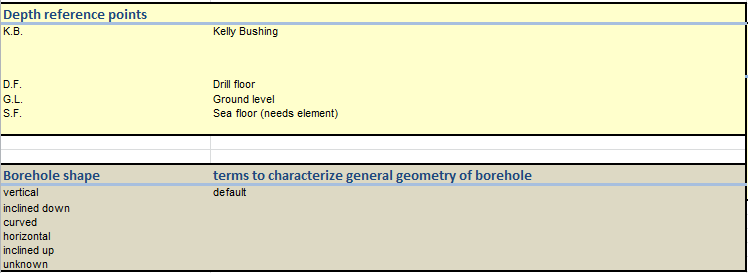


Figure : A sample entry from the Data Valid Terms spreadsheet

# ENTERING DATA

To enter data, click the Template spreadsheet and copy your records into the spreadsheet, field by field.

For example, assume that you have information about eight well boreholes. Your information includes the following:

* An API number for each borehole
* The owner and operator of the well associated with each borehole
* The latitude/longitude coordinates for the well associated with each borehole

As indicated by the table at the bottom of this page, this information should be entered into the WellHeader template. Each row should constitute an individual database record, so including the top row (which contains field information), there should be n total rows of data in the template.

Data should then be entered into the corresponding column (Table 6). Data regarding data entry can be found on the Field List tab.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **APINo** | **Operator** | **MineralOwner** | **LeaseName** | **LatDegree** | **LongDegree** |
| 02-021-05000 | Nichols, Hugh E. | State | State | 33.33789906 | -111.5474776 |
| 02-005-05019 | Sinclair Oil & Gas | Santa Fe Pacific RR | Santa Fe | 35.76820705 | -112.3573439 |
| 02-001-05335 | Kerr-McGee | Hortenstine-Macie | Macie Fee | 35.08278252 | -109.5836642 |
| 02-003-05038 | Allen, Elmer R. | W. J. Davis | Davis | 31.57815012 | -109.7510095 |
| 02-017-05065 | Texaco-Skelly-Sinclair | Navajo | Navajo | 36.98885669 | -110.3807293 |
| 02-021-05005 | Western Oil Fields | Federal | Federal | 32.94538466 | -111.3689436 |
| 02-027-05011 | Stewart, M. P. | Federal | Federal | 32.52173669 | -114.6501276 |
| 02-001-05198 | Shell Oil | Navajo | Navajo | 36.98311134 | -109.2666708 |

Table : Sample entries in a template spreadsheet

It should be noted that some datasets in a larger number of fields than is accommodated by the content model. Using the example from Table 6, assume that you also wish to submit data regarding temperature observations within all eight boreholes, as well as detailed information about the metallurgical properties of the drill bits used to drill the borehole. Under these circumstances, you have three options:

* Separate the data into multiple appropriate content models
* Omit data that is not accommodated by the content models
* Serve the data yourself

SGD possesses a separate content model designed to accommodate borehole temperature observations. Consequently, it would be possible for you to enter information about the boreholes themselves in the WellHeader content model and information about borehole temperature observations in another.

Unfortunately, SGD does not possess a content model designed to accommodate the metallurgical properties of boreholes. In this case, you may choose to omit the data that is not accommodated by any content models, or serve the data using your own resources. If you omit data that is not accommodated by any content, this may cause the omitted data to remain undiscovered and unused. On the other hand, if you choose to serve your own data, you are wholly responsible for its upkeep.

# SUBMITTING DATA

After you have prepared your data for submission by entering it into the template spreadsheet of a content model, you may submit your data through the State Geothermal Data website.

* Login: Click **Login** in the upper right hand corner of the SGD website (Figure 11). Enter your username and password in the login window that appears. Alternatively, you may login via established providers such a Gmail, AOL, and OpenID. To do so, simply click the appropriate icon at the bottom of the login window and enter the username and password you normally use for that provider; your provider will then log you in to SGD.
* Submit your data: Mouse over the **Members** menu and click **Submit** Files (Figure 12). This will bring you a Submission page (Figure 13) on which you will provide detailed information about the data you wish to submit. Required fields are indicated by a red asterisk (\*). Click the **Save** button, or click **Preview** to see what your data will look like when submitted.
* Wait for feedback: Someone at the Arizona Geological Survey will review your data and let you know whether or not it needs to be revised. If your data needs to be revised, you may resubmit it for review.

If your data is approved and does not need to be revised, the following steps need to be taken:

1. Provide the AZGS with a metadata record for your data using the USGIN Metadata Wizard. If your data is being served as a service, you will need to provide the AZGS with a metadata record for the service, as well as for the data.
2. Get your data online. If your data is a deliverable (for example, a completed content model), it will need to be uploaded to the USGIN document repository. If your data is going to be hosted as a service, this can be accomplished in one of the following ways:
   1. If you are hosting your own data, you will need to mobilize sufficient computing resources
   2. If you are a California or Arizona subrecipient, AZGS will host your data as a service
   3. If you are a subrecipient in any other state, AZGS will notify your state of the fact that your data has been approved, and they will host your data as a service.



Figure : Logging in to SDG

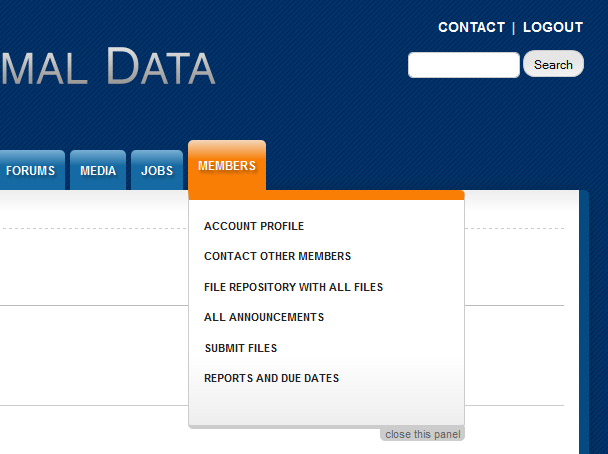


Figure : The Members menu

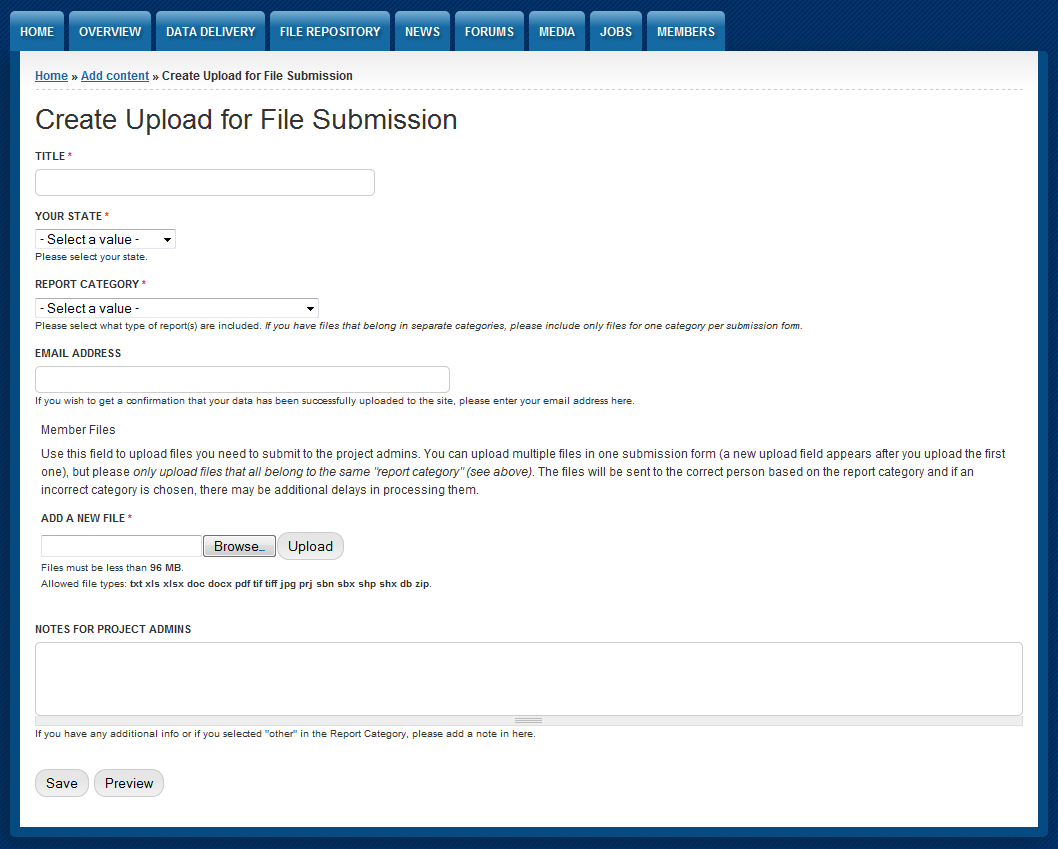


Figure : The Submission page

# SCHEMAS, CONTENT MODELS, TEMPLATES, AND INTEROPERABILITY: AN EXAMPLE

The following example will demonstrate the importance of content models and templates in facilitating interoperability at the data level.

Two ornithologists are recording data about birds. Tables 3 and 4 contain their results.

|  |  |  |  |
| --- | --- | --- | --- |
| **Specimen ID** | **Species** | **Wingspan** | **Bill Length** |
| 2312 | *C. Cristata* | 34 cm | 2.2 cm |
| 2313 | *C. Cristata* | 35 cm | 2.5 cm |
| 2314 | *C. Cristata* | 37 cm | 2.3 cm |
| 2315 | *C. Cristata* | 34.5 cm | 2.2 cm |
| 2316 | *C. Cristata* | 36 cm | 2.1 cm |
| 2317 | *C. Cristata* | 36 cm | 2.4 cm |

Table : Blue jay sample data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SpecimenURI** | **Sp.** | **Wingspan** | **Beak Length** | **Units** |
| j:12 | *Larus argentatus* | 48 | 1.1 | In |
| j:13 | *Larus argentatus* | 55 | 1 | In |
| j:14 | *Larus argentatus* | 54 | 1.15 | In |
| j:15 | *Larus argentatus* | 49 | 1.2 | In |
| j:16 | *Larus argentatus* | 59 | 1.12 | In |

Table : Seagull sample data

Both tables obviously record the same kind of data, which matches different individuals of a given bird species with corresponding measurements of beak and wingspan.

But these tables are not interoperable because they use different schemas. In Table 3, specimens are listed under the **Specimen ID** field; in Table 4, specimens are listed under the **SpecimenURI** field. In Table 3, **Species** is spelled out, but the genus is abbreviated; in Table 4, **Species** is abbreviated as **Sp.** but genus is spelled out in each entry. Likewise: in Table 3, measurement units for specimen wingspan and beak measurements are provided in the same field as the measurements themselves; in Table 4, the units for these measurements are listed in a separate field. Finally: in Table 3, beak measurements are listed under the **Bill Length** field; in Table 4, they are listed under the **Beak Length** field.

All of this might not seem like much to a human operator, who is able to understand and account for the discrepancies between the two tables. But for a computer, such intellectual leaps are impossible. Likewise, there are occasions in which schematic differences such as those above can baffle even human operators.

In order to make data sharing as easy as possible, SGD uses content models to ensure that data providers submit their data the same way. This is not to say that the content models used by SGD are somehow “correct.” Rather, the content models employed by SGD represent a practical compromise between machine readability, human readability, and the demands of the data.

Now imagine that the Ornithology Society has decided to adopt the following schema for data about birds:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Specimen** | **Species** | **Wingspan** | **Beak/Bill Length** | **Measurement Units** |
| Unique identifier for each bird measured | Genus and species of specimen; provide full names for both genus and species | Measurement of specimen from wingtip to wingtip | Measurement of specimen's beak or bill length | Abbreviated units in which all measurements for a specimen are recorded |

Table : A content model for an ornithological schema

When fitted to the above schema, the data from Table 1 and Table 2 would appear as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Specimen** | **Species** | **Wingspan** | **Beak/Bill Length** | **Measurement Units** |
| 2312 | *Cyanocitta cristata* | 34 | 2.2 | cm |
| 2313 | *Cyanocitta cristata* | 35 | 2.5 | cm |
| 2314 | *Cyanocitta cristata* | 37 | 2.3 | cm |
| 2315 | *Cyanocitta cristata* | 34.5 | 2.2 | cm |
| 2316 | *Cyanocitta cristata* | 36 | 2.1 | cm |
| 2317 | *Cyanocitta cristata* | 36 | 2.4 | cm |
| j:12 | *Larus argentatus* | 48 | 1.1 | in |
| j:13 | *Larus argentatus* | 55 | 1 | in |
| j:14 | *Larus argentatus* | 54 | 1.15 | in |
| j:15 | *Larus argentatus* | 49 | 1.2 | in |
| j:16 | *Larus argentatus* | 59 | 1.12 | in |

Table : Data from Table 1 and Table 2 consolidated under a unified schema

Since their data now uses the same schema, the data from Tables 3 and 4 is now interoperable and can be directly compared without operator interpretation.

Note that the measurements for *C. Cristata* are still in centimeters and the measurements for *L. Argentatus* are still in inches. Though this complicates a direct comparison between the two species, it does not prevent interoperability between the two data sets: they are still in the same schema, and the schema allows for measurements of different units. The schema also makes it very easy to search the database for all records measured in centimeters or inches, since the measurement units are contained within their own field. No operator interpretation is required.

Note also that the specimen identifiers use different formats. Though it might countermand best practice, this does not prevent interoperability either, as long as each identifier is unique.