# Contents

Document Versions......................................................................................................................................................... 5

Introduction........................................................................................................................................................................ 6

FGDC Metadata .................................................................................................................................................................. 6

Metavist ............................................................................................................................................................................. 7

Get Started ..................................................................................................................................................................... 8

Identification ................................................................................................................................................................... 9

Basic Info ..................................................................................................................................................................... 9

   Primary Citation ....................................................................................................................................................... 10

   Larger Work Citation ............................................................................................................................................... 12

Basic Information Completion ......................................................................................................................................... 14

Spatial Domain ............................................................................................................................................................... 15

Keywords ....................................................................................................................................................................... 15

Taxonomy ....................................................................................................................................................................... 17

   Classification System or Authority ......................................................................................................................... 19

   Voucher .................................................................................................................................................................... 20

Access .......................................................................................................................................................................... 21

   Point of Contact ....................................................................................................................................................... 22

Analytical Tools ............................................................................................................................................................ 22

Miscellaneous .............................................................................................................................................................. 23

   Cross Reference ....................................................................................................................................................... 24

Data Quality .................................................................................................................................................................. 25

   Attribute Accuracy .................................................................................................................................................... 25

   Consistency & Completeness .................................................................................................................................. 25

   Positional Accuracy ............................................................................................................................................... 26

Lineage .......................................................................................................................................................................... 27

   Methodology .......................................................................................................................................................... 28

   Source Information ................................................................................................................................................. 28

   Process Steps ........................................................................................................................................................ 28

Cloud Cover ................................................................................................................................................................. 30

Spatial Data Org............................................................................................................................................................. 31

   Reference Methods ............................................................................................................................................. 31

   Vector Information ............................................................................................................................................... 32

   Raster Information ................................................................................................................................................ 33

Spatial Reference .......................................................................................................................................................... 34

   Horizontal Coordinate System Definition – Geographic Example ........................................................................ 34

   Horizontal Coordinate System Definition – Planar Example ............................................................................. 35
Figures
Figure 1: Opening screen ................................................................. 8
Figure 2: Basic Info screen .............................................................. 10
Figure 3: Primary Citation screen .................................................... 11
Figure 4: Primary Citation completed ............................................ 12
Figure 5: Larger Work Citation screen ........................................... 13
Figure 6: Basic Info completed ...................................................... 14
Figure 7: Spatial Domain screen .................................................... 15
Figure 8: Keywords screen ............................................................ 16
Figure 9: Theme Keywords list screen ......................................... 18
Figure 10: Taxonomy screen ........................................................... 19
Document Versions

1.1: (8/18/16) Draft for peer review.

1.2: (8/26/16) First release.

1.3: (12/1/16) Revisions to Positional Accuracy Report section.

1.4: (1/27/17) Revisions to Primary Citation section, adding ORCID and suggested citation; revision to Entity and Attribute section to describe how to change an attribute domain type.
Introduction

The following is an unofficial general guide to using the Metavist metadata editor application. The intent of this guide is to help users with understanding the interface and to provide guidance on what content should be provided in each form field. Following this guide will help in developing a robust and valid metadata record that follows general Alaska Science Center (ASC) Data Management best practices. As always your mileage may vary. There is a great variation in data sets and addressing that complexity is beyond the scope of this document. Other editors and tools are available that may work better for your situation or in combination with Metavist. Please see Metadata Resources on the ASC Data Management website for a list of tools and information: http://ascinternal.wr.usgs.gov/datamgt/metadata_resources.pdf

FGDC Metadata

Metadata is documentation of data in a machine readable, structured format. While metadata is primarily used to document data, it is also used to document software, projects, models, collections, maps and documents. We will use the term “data resource” in this document to refer to this broader definition. A metadata editor creates an XML formatted file that consists of pre-determined tag sets defining a structural framework for the content. Discrete information can be read by applications for various purposes by interpreting these tags. The meaning and structure of the XML tagging is defined through a standards specification. There are a number of metadata standards, but we are utilizing the Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (FGDC CSDGM) or commonly referred to as the FGDC standard. This is the current U.S. standard. In addition, there are extensions to the core standard called profiles, namely, Remote Sensing, Shoreline and Biological. Since environmental sciences are a significant research area of the ASC, we utilize the Biological Profile in addition to the core standard. This extension provides additional elements for taxonomy and methods and will be discussed further in this document. The FGDC standard has been deprecated to be replaced in the near future by the international ISO 19115 standard. We are preparing for migration to ISO, but at this time, FGDC is still the recommended standard. This document will in places try to explain expected content by the FGDC standard. However, please see the FGDC documentation for more information. Links are available from Metadata Resources:

FGDC CSDGM Data Standard Document: http://www.fgdc.gov/metadata/csdgm/

FGDC Metadata Map GUI Tool: http://www.fgdc.gov/csdgmgraphical/index.htm

USGS FGDC Data Standard Extensions Listing: http://www.usgs.gov/core_science_systems/csas/metadata/standards.html
Metavist

Metavist is a freeware application developed by David Rugg, U.S. Forest Service. It is a lightweight, stand-alone application developed specifically to run on Windows. It fully supports the FGDC core data standard as well as the biological profile. While it works very well, the interface can be labyrinth-like in places. You will need IT to download and install the code. If you go out on your own to acquire the software, please be aware of the availability of a 2.0 version. That version is said to be in beta, but it is an essentially never finished and abandoned project. This buggy version will corrupt your files, as such should be avoided. Use the 2005 version available directly from the Forest Service. Both the application and documentation by the author are available on the FS website: http://www.nrs.fs.fed.us/pubs/2737. The documentation will provide further background on Metavist and the FGDC standard and is recommended for general understanding of the interface.
Get Started

To get started, open the app. The opening screen will display with three buttons. Hovering over buttons throughout the app will provide tool tips. The first button on the opening screen will start a new file, the second will open an existing file, the third is the save function. These functions are also available from the menu under File.

IMPORTANT: Metavist is unforgiving about saving your work. Make sure you click to save before exiting or all your work will be lost. No helpful reminder or incremental save feature here. Best to save now and then as a precaution. Saving after each tab entry is a good safe practice.

Strategically, it is usually best to start with a template to speed entry and ensure standard entry for commonly used fields. There are a number of templates started by various offices, or see a data manager (DM) for help. Another recommendation is to review the ASC products listing for relevant examples of metadata:

http://alaska.usgs.gov/products/data_all.php

Figure 1: Opening screen
Identification

The user interface uses general tabs, then sub-tabs under each general tab. Identification is the first tab. We will discuss each sub-tab under it, which will be referenced as a screen.

Basic Info

Once you start a new document or open an existing one, the Basic Info screen will appear. The first section on the screen is the primary citation followed by dates and maintenance sections. Click on “Edit Citation” to create the primary citation.

The FGDC standard has elements that are considered mandatory, mandatory if applicable, and optional. Metavist more or less tries to indicate elements that are mandatory or mandatory if applicable, by using a bold font for the title. Optional elements are indicated with non-bolded names and possibly an optional indicator. Further complicating the designation, ASC best practices may consider an optional element to be mandatory. These cases are identified in this document. An XML file is run through “validation” to verify it meets the minimum requirements of the FGDC standard and is a valid XML formatted file. Data management uses the USGS Metadata Parser for this purpose. See Metadata Resources for access to the validator and see the FGDC documentation for more information on what is considered mandatory by the standard.

USGS FGDC Metadata Standard Validator Tool:
http://geo-nsdi.er.usgs.gov/validation/
Primary Citation

The Primary Citation screen will popup. This is the entry screen for the data resource title and citation for the creators of the data resource. Note: some try to make the metadata citation match with a publication. The data resource is a product in its own right and the metadata is documentation for the data resource, thus the citation should be from that perspective. The title is for the data resource as well and is what appears in various metadata catalog data listings. The recommended format for title is data resource name, general geographic area, year range, as is applicable. Ex: Migratory Bird Avian Influenza Sampling; Yukon Kuskokwim Delta, Alaska, 2015. Authors are intended to be the authors of the data resource, not necessarily authors in a publication. Collaborators and contributors to the data resource can be referenced in the Data Set Credit section. The author ORCID should be provide after the author’s name in the format: Powell, J.W. (ORCID xxxxxxxxx).
Figure 3: Primary Citation screen

Data set creators, one per each line, no separators. Include ORCID after author.

Usually left blank.

Anchorage, AK | U.S. Geological Survey, Alaska Science Center

Suggested citation for the data set.

DOI for data release. DM will add this.

Click “Yes” to open larger work citation screen.

Ok to save, cancel to drop edits; close screen.

Estimate of data release: month and year are adequate.

Data set title: name; geographic area, year range.

File format of data.
Figure 4: Primary Citation completed

**Larger Work Citation**

This screen is another citation for an umbrella collection that the data resource belongs to. We use this to reference the project in the USGS ASC Science Portal: [http://alaska.usgs.gov/portal/project.php?project_id=18](http://alaska.usgs.gov/portal/project.php?project_id=18). By doing so we provide a reference to a general description of the study the data resource was collected for. It is accessed through the Primary Citation screen by clicking on the Yes radio button next to “Display Larger Work Citation?”. The following is a completed screen and is largely self explanatory since it contains the same fields as Primary Citation. Content for this screen can be obtained from the project record in the Portal. Be sure to click on OK to save your entry.
Figure 5: Larger Work Citation screen

Science Portal project record URL.
Basic Information Completion

After returning from the Primary Citation screen, you will see some of the information in the Basic Information screen greyed out. To edit, you must return to the Primary Citation screen and complete the rest of the fields. The abstract field content should be a succinct paragraph that provides a general explanation of what the data resource is. This should not be the study or publication abstract and should be devoid of any interpretation, analysis, or results of your research. Think of it like a movie listing. It provides an at-a-glance description to allow a viewer to quickly assess the data’s potential use. Purpose is a succinct statement that describes the purpose for which the data resource was collected relative to the study, it is the “why” companion to the abstract’s “what”. For example, “To examine the genetic relationships among Larus gulls using samples from North America, Europe, Russia, and South America”.

Figure 6: Basic Info completed
Spatial Domain

This screen is used to describe general geographic areas the data resource pertains to. Bounding Coordinates are considered mandatory and supports geospatial discovery in metadata catalogs. There are occasions when this is not applicable, such as a software resource, but should otherwise be provided. A useful web tool for determining bounding coordinates is here: http://boundingbox.klokantech.com/

![Figure 7: Spatial Domain screen](image)

Keywords

Keywords are controlled vocabularies managed in thesauri by various organizations. They are used to enable word searches in metadata catalogs. Use thesauri as relevant to your subject area. There are a few that we use at ASC as standards where applicable, please see Metadata Resources for access to thesauri:

<table>
<thead>
<tr>
<th>ISO Topic Category</th>
<th>At least one is required for compatibility with the ISO standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA GCMD Earth Science</td>
<td>Covers many subject areas, should have at least one</td>
</tr>
<tr>
<td>USGS Biocomplexity</td>
<td>Also covers many subject areas, should have at least one</td>
</tr>
<tr>
<td>USGS GNIS</td>
<td>Required, specific for geographic names</td>
</tr>
<tr>
<td>NGIA GeoNames</td>
<td>Foreign geographic names, as applicable</td>
</tr>
<tr>
<td>MeSH</td>
<td>Optional, good for genetics keywords</td>
</tr>
</tbody>
</table>

In addition, you can create an unnamed thesaurus keyword list for keywords not covered by a controlled vocabulary. Use “None” for thesaurus name.
To create a keyword list, click Add. Once you complete the list and save, it will show in the window. Click on it again to edit the list or to delete. Click on Add to create additional keyword lists. Note, each keyword list is specific to a thesaurus.

After clicking Add or Edit, the Keyword list screen will popup. Enter the thesaurus. Try to use standard nomenclature. Note: For NASA GCMD, the keywords come in a hierarchy: Earth Science > Biological Classification > Animals/Vertebrates > Birds. Separate each level into a separate keyword as seen below:
**Taxonomy**

Required as a part of the FGDC standard Biological Profile extension, taxonomy describes the taxonomic classification for species referenced in the data resource and provides documentation for how the species were identified. It is mandatory for biological subject areas. The profile assumes use of the Integrated Taxonomic Information System (ITIS available at: [http://www.itis.gov/](http://www.itis.gov/)) as a default. However, other classification systems may be used. Example given is for ITIS. Taxonomy is effectively another type of keyword supporting data discovery by species, genus, etc. Taxonomic classification can be created utilizing other tools.

The taxonomic classification is a tree structure. For each level you specify rank name and value and optionally common name. To create a level, use Add Child. You can add species for multiple genus, order etc. by simply adding a child at the requisite parent level. Note: there is no validation or entry constraint on rank name, so it is up to you to get the ranking order correct.

The Import button is intended to import an existing classification snippet from an XML file into this section, though it’s not as straightforward as it appears. See Metadata Resources for further procedural information on this function.

Please see the Biological Data Profile document for a more detailed explanation of field descriptions: [https://www.fgdc.gov/standards/projects/metadata/biometadata/biodatap.pdf](https://www.fgdc.gov/standards/projects/metadata/biometadata/biodatap.pdf)
While optional, Taxonomic Coverage is a useful place to give an overview of species. For example: “marine birds in the Gulf of Alaska or freshwater invertebrates from inland waterways in Alaska.”
**Classification System or Authority**

Identify the classification system used for taxonomic identification.

Provide a citation for the taxonomic system. In many cases the ITIS TSN number can be provided for species, genus, order, etc. In other cases with multiple species across genus it may be impractical. The TSN number and URL link to the ITIS classification report can be obtained after locating the taxonomy for your species of interest.
Figure 12: Classification System Citation screen - ITIS example

Voucher

Citations for repositories holding the samples used.

Figure 13: Voucher screen

ITIS identification number

Use this format, replace with actual retrieval date from ITIS.

Link to ITIS classification report.

Specimen type. If suitable value not available, type in your own.
Access

This screen is used to describe any constraints on access or use of the data and provides a point of contact.

Generally, we cannot constrain public access to data. Enter “None”.

Not used.

Generally, we cannot constrain public usage of the data. However you can ask for citation.

Point of contact.
**Point of Contact**

Primary Point of Contact while optional is recommended. We use a generic POC and ASC phone and email. This is due to the likelihood of personal information going stale over time. Secondly, public questions may be about access issues which are more appropriate for IT or perhaps DM. Though not recommended, you can specify a name for the POC, just make sure to use the ASC phone and email.

![Image of Primary Contact screen](image)

**Analytical Tools**

This screen is used to describe tools, models, or statistical procedures that a data resource is intrinsically bound to and available to the public for use with the data resource.

Enough information should be included so that a potential data user can easily determine why they might wish to acquire the analytical tool, and instructions on how to acquire it.
**Miscellaneous**

In spite of the name there are a couple of important sections we like to use.

Data Set Credit is an optional field that can be used to identify collaborators, researchers, partner or funding organizations, etc. who had a contributory, indirect role in creation of the dataset. Some like to cite ASC. While not wrong, ASC is already cited in the primary citation and this field is intended to recognize additional contributors not already cited.

Native Data Set is an optional field that can be used to identify the processing environment for the data resource, such as OS, Software and version.

Browse Graphic is an optional section that can be used to link thumbnail images to the metadata for use by metadata catalogs as a graphic in a listing. We do not recommend this as we are not setup for serving these images at this time from the ASC portal and any links would likely be fragile.
Cross Reference

Cross Reference is an important section and can be used to provide a citation for a relevant publication(s), resources, project and program web sites, collaborator sites, etc. To enable the ASC products listing, we ask that the first citation be for the publication a data resource being release is associated with.

Figure 18: Cross Reference Citation screen - publication example
Data Quality

This tab is for providing information about how the data was collected and processed, identification of significant field or lab methods applied, how QA/ QC was done and citation for any existing data sources used in development of the data resource.

Attribute Accuracy

Attributes are “columns” in which discrete data is stored, such as, id, species code, location, etc. In this text field, provide a description of how the accuracy of value assignments were determined. For example, “The assignement of survey information to respective fire category attributes were accomlished using automated R scripts to minimize manual errors.” Each attribute should be addressed where reasonable.

Consistency & Completeness

Logical Consistency is intended as a place for summarizing things, such as, topological checks for geospatial, QA/ QC evaluations of values within a range, identification of a text value used within a numeric field such as an “NA” or “Not Applicable” or meaning of a blank value. For example: “Ten percent of the samples from the breeding populations were extracted, amplified, and genotyped or sequenced in duplicate for quality control.”
Completeness Report is the identification of data omitted from the data resource that might normally be expected and reasons for the exclusions. These may include outliers, misidentifications, depleted samples, etc. For example: “A blank cell signifies that the test was not run for this sample.”

**Positional Accuracy**

Describes the accuracy of any positional data in the data resource. If there are no positional data than this section can be ignored. Separately describe horizontal accuracy and if applicable, vertical accuracy. This can be in a narrative format and/or quantitatively. Horizontal accuracy should be provided for geospatial data resources. Identifying methods for how positions were acquired should be specified. For instance: “estimated position from Google Earth at an eye altitude of 16 miles,” or “positions acquired using a Garmin hand held GPS with an estimated 3 m accuracy.”

![Figure 20: Positional Accuracy screen](image)

Note: There is a glitch in the application in regards to Vertical Positional Accuracy Report. If entered once and saved it will be saved as expected. However, if the record is
re-edited and saved again FOR ANY PURPOSE, Metavist will strip this section out. It will simply disappear after the save. Therefore it is recommended this information be recorded in other places, such as within the Horizontal Positional Accuracy Report.

**Lineage**

This screen is used to describe the methods, process steps and provide citations for any existing data resources that were used in developing the data resource. This information allows a user to document how the data resource was derived and what protocols or methods were followed. Process steps and source have applicability for software. The interface section works similar to other sections you have seen in that you add something to a list using Add, or select an item to edit or delete. In the case of add and edit, a popup screen will appear for entry of relevant information.

![Figure 21: Lineage screen](image)
Methodology

If you are using standard protocols than this is a good use of this section to provide reference to them. You can just provide a general statement and a citation to the publication that describes the method. Required if using the Biological Profile.

![Methodology screen](image)

**Source Information**

Source information is often overlooked but it is important to cite existing resources used in developing your data resource. This is particularly the case with geospatial data in which a remotely sensed product, DEM, low altitude image, etc. was used. This can also be used to identify the model used for an output data source or to describe process function, inputs/outputs and other factors of software.
Citation for the source.

Description of how the source was used in this dataset.

Identifier for the source.

Optional, but applicable for map source.

Should provide temporal reference for the source.

Figure 23: Source Information screen - map example

Figure 24: Citation screen for a Source
**Process Steps**

Process step often get confused with Methodology. Process differs from Methodology in that it identifies a series of discrete steps taken to derive the resource rather than a listing of standard methods applicable to the resource, as is provided in Methodology. Process Steps have particular significance for geospatial data and software, but also has applicability for many other data resources as well.

![Process Step screen](example.png)

**Cloud Cover**

Cloud cover is the last screen within Data Quality and is only applicable to aerial or satellite imagery. It is intended to record % image obscured by cloud cover. If not applicable, leave as “Not reported”.

Optional fields, rarely used.

General description of the process step

Date associated with the process step.
**Spatial Data Org**

This tab is used to provide information about how geospatial data is represented in the data resource. Even if you have no geospatial data in your data resource, you may have approximated geographic references. If you have neither geospatial nor named geographic features, then this section can be ignored.

**Reference Methods**

In this screen you specify how geospatial locations are represented.

Indirect Spatial Reference is used to specify geographic locations by references, such as geographic features, named locations, addressing schemes, etc. This is not for providing a list of geographic names, but to describe the type of referencing used. This field may be applicable for non-geospatial data or in conjunction with Direct Spatial Reference Method for geospatial data with named geographic references.

Direct Spatial Reference Method is a two type classification for geospatial data. The first level is whether the data is vector or raster. Secondly, if vector whether the features are point or areal/linear. This field is mandatory for any geospatial data. If the data is not geospatial, set to blank.

![Spatial Data Org screen](image_url)
Vector Information

Vector Information is intended by the standard to be filled out if you choose Point or Vector Reference Methods. It is essentially used to list the feature and topological classes used in your data resource. There are two standards for this description: Spatial Data Transfer Standard (SDTS) which is an FGDC standard, or Vector Product Format (VPF) which is a Dept. of Defense standard. No description is also an option. Using SDTS terms is the likely usage. Once picked, you can create a list of feature classes. If you are using ArcGIS or other geospatial processing application, it is sometimes helpful to start a metadata record in that system, which will provide this information automatically. Generally speaking I think this screen is getting into the weeds as far as providing useful documentation. “No description” is a perfectly valid choice.

Figure 27: Vector Information screen

Figure 28: SDTS Terms Description screen - point example
**Raster Information**

This screen is only used if you picked Raster for the Reference Method. This is a useful section to identify the grid layout for some types of raster data.

![Figure 29: Raster Information screen](image)

- **Type of raster.**
- **Dimensions of grid.**
**Spatial Reference**

This tab is used to describe the geospatial reference system associated with geospatial data. If you don’t have geospatial data, check this section anyway to make sure it is not creating one, as many templates have a reference system set as a default. Note, the term geospatial does not just apply to a data set created in a geospatial application. If you have coordinates in your data set, you have geospatial data.

If you are using ArcGIS or other geospatial processing application, it is sometimes helpful to start a metadata record in that system, which will provide this information automatically.

**Horizontal Coordinate System Definition – Geographic Example**

This section describes the horizontal reference system. This section is required if you have geospatial data. If you do not have geospatial data, make sure the Horizontal Coordinate System Definition is set to “No definition”.

Geographic is applicable if you have latitude and longitude values in your data set. Choose “Geographic” if appropriate. This will change the accessibility of fields in the screen to what is appropriate for a geographic specification.

In the Geographic box you set the lat and long coordinate precision and the coordinate units.

In the Geodetic Model you will identify the datum, the ellipsoid and definition parameters. Many are using a WGS84 datum and ellipsoid. Note that the Semi-major Axis is the same for both GRS80 and WGS84 ellipsoids. However the Denominator of the Flattenning Ratio is slightly different for WGS84. Metavist defaults parameters for GRS80 which is 298.257222101, so you may need to adjust the value for WGS84 to: 298.257223563.

Planar and Local boxes are not used for geographic.
Horizontal Coordinate System Definition – Planar Example

If your coordinate system is projected, then you will set your definition to “Planar”. This will change the fields access accordingly such that you will use the Planar section rather than the Geographic section. This is a more complicated section as you need to set the projection parameters.
Coordinate System (UTM)

In this screen you define your projection and coordinate system. For UTM it is a little confusing. Select the Grid Coordinate System option. This will allow you to set the coordinate system to UTM and the projection will be auto-filled to a Transverse Mercator. You will need to click on Edit Parameters to set the parameters for the UTM zone.

Figure 32: Planar Coordinate System Definition screen - UTM example
Planar Parameters (UTM)

For the checked box parameters you will need to provide the parameters for the UTM zone. Click OK to return to the Coordinate System screen. Zones can be determined by a number of online references such as this: http://www.jaworski.ca/utmzones.htm
**Planar Coordinate Information**

In this screen you specify coordinate system units and resolution. The resolution is intended to define the coordinate granularity to represent the features in the data resource.

![Planar Coordinate Information screen](image)

*Figure 34: Planar Coordinate Information screen*
**Vertical Coordinate System Definition**

This screen is used to define the coordinate reference system for elevation or depth coordinates.

![Vertical coordinate system definition example](image)

- **Vertical Datum.**
- **Set coordinate precision, or list if variable.**
- **Units of measure.**
- **How elevation coordinates are included.**

Figure 35: Vertical Coordinate System Definition - elevation example
**Entity & Attribute**

This tab provides a data dictionary for the data resource in which you will account for the entities and attributes, their definitions and domains. A data dictionary fully describes these data elements such that subsequent users can understand what was collected.

Entities are physical containers for data, such as spreadsheets, database tables, sensor log, etc. Entities have definitions and attributes. Attributes are the “columns” in which discrete data is stored, such as, id, species code, location, etc. Attributes have definitions and domains. Domains are the set of possible/observed values for the attribute. For example, M, F for Sex; 0-32 degrees C for Temperature, etc.

Some create the data dictionary externally, such as using an XML editor. When working directly on an XML file, make sure you know the standard, keep the tags and structural integrity intact and make a backup before beginning. It is a good idea to verify structural integrity of the XML by opening the file in Metavist and validation using the metadata parser, or utilize an XML stylesheet (XSD). Others may author the data dictionary as a Word document then copy/paste definitions into the Metavist text fields.

A dictionary can be documented in Metavist using either an Overview Description, a Detailed Description or both as relative to the data resource. In some cases a data dictionary is not applicable, such as a software resource.

It is best to finalize the attribute list and ordering within your data set before creating a data dictionary as ordering cannot be edited easily within Metavist.

**Overview Descriptions**

Jumping to the second sub-tab, this section is occasionally applicable. For instance, if you are documenting a proprietary format file, it may be more appropriate to provide a reference for an external data dictionary defined by the proprietor. Note that in this screen you are creating a list of overview descriptions, so you can have multiple references.
Overview Description

In this screen you can define an instance of an overview description.
**Detailed Descriptions**

Detailed Description should be the default in most circumstances in which the user creates a data dictionary specific for the data set.

In the Detailed Descriptions screen, entity, attribute and domain sections each have a drop down list control. When clicked you will see the list of elements you have created so far using Add. The list is in the order they were created. To edit or delete an existing element, use the drop down list to select the appropriate element. Make sure it now appears in the box window. Edit and Delete will operate with this selected element. Verify that the element you wish to delete shows in the box before pressing the button as it is very easy to make a mistake here.

**Detailed Descriptions Interface**

There are two aspects to creating a Detailed Descriptions data dictionary. One aspect is to define each structural element of the data dictionary: entities, attributes and domains. The second aspect is to delineate the element’s inter-relationships by attaching the elements together as hierarchical “branches” (entity->attributes->domains). First we will discuss creation of each of the elements then discuss relating elements together. Looing at a completed sample metadata file on the ASC Data Products listing with provide a visual depiction of how this hierarchy works.
It may be advantageous to develop a data dictionary document (such as Word) for your data set before hand. Then copy/paste definitions from the document into the appropriate Metavist text fields. Many offices have developed standardized data dictionaries for their subject areas providing standardized definitions for attributes and domains. This saves a great deal of time for commonly used elements, provides consistency and improves quality. Peer review through usage also increases quality. Check with data stewards, branch chief or a data manager for more information on existing data dictionaries.

**Entity Type**

This screen is used to provide the entity name and a definition. In many cases, such as singular spread sheet data sets, the entity is the spread sheet and the name of the entity is the name of the spreadsheet.

![Entity Type Description screen](image)

- **Name of the entity.** Usually spread sheet name or table name. File extensions are unnecessary.
- **Entity definition.** Definitions should be complete sentences using clear language. You can also copy/paste from other sources.
- **If using an externally defined definition, cite the source.** If using your own then specify as “User defined” or “Author defined”.

Figure 39: Entity Type Description screen
**Attribute**

This screen is used to provide the attribute name and a definition.

![Attribute screen](image)

This will list in the drop down whatever domains you have attached to this attribute through the posting tools.

This will list whatever domains you have attached to this attribute through the posting tools.

**Domain Type**

This screen creates a domain. A domain can be attached to multiple attributes, and an attribute can have more than one domain, but must have at least one domain. There are four types of domains; enumerated, range, codeset and unrepresentable. Depending on what domain type you choose, the appropriate section in the screen will become active.

**Enumerated:** A discrete code value used in the dataset. Only values actually used in the dataset should be listed. For example: an attribute of “species” will have an enumerated domain for the name of each species in the data set.

**Range:** Identifies min and max values for the attribute. For example, an attribute of “sea ice thickness” will have a range domain defining the min and max values occurring in that attribute.

**Codeset:** Essentially a citation for an externally accessed domain. Generally, specific values used in the dataset should be enumerated. In some cases, enumeration is impractical and in that case a reference can be made.

Precision of measured values and explanation (if any) regarding expressed or implied accuracy.
Unrepresentable: This should only be used for cases where a defined domain is not relevant or impractical to define. Describe why the domain cannot be represented. Again, you must identify a domain, but can use Unrepresentable if one cannot be defined. For example, GenBank accession number or museum voucher number for a specimen. In both cases numbers are not ordered sequentially or have minimum and maximum values that make sense.

Choose the domain type.

Figure 41: Attribute Domain Values screen
Depending on what domain type you choose, the appropriate section in the screen will become active.

![Figure 42: Attribute Domain Values - range domain example](image)

Enumerated values are a domain instance consisting of one enumerated value. In essence you create a “domain” for each value in the domain. For instance, a country domain consisting of France, Switzerland and Germany would have three separate domain entries, one for each country. Each value will be attached individually to the applicable attribute to comprise the complete domain.

You can change the domain type while editing, but once saved, the domain type is set and cannot be changed. Instead, the old domain would need to be deleted and a new one created with the desired domain type.
Attaching Elements

The approach in Metavist is to create a series of hierarchical branches. Starting with entities and attaching “branches” of attributes to entities. In this way you are defining which attributes occur with which entity and the order of that occurrence. An attribute can be attached to more than one entity. An entity can have as many attributes as you like. The attribute naming and ordering should match your data resource.

The next level of branching are domains. Domains are attached to attributes. Domains can be attached to more than one attribute and attributes can have multiple domains.

Though you can vary when you attach dictionary elements together relative to your work flow, there is a caveat. Any unattached elements will NOT be saved despite a file save. You must complete the attaching sequence before closing out your session.

See the ASC Data Products listing for visual examples of this hierarchy.
**Attaching an Attribute to an Entity**

To attach an attribute to an entity, thus creating a branch, you will return to the Detailed Description screen. Use the following steps:

1. Select the entity to be the “root” for the branch from the entity drop down list. Make sure it is visible in the window.
2. Click on the arrow tool to the right of the entity drop down list. The entity name should show in the Root box.
3. Select the attribute to be the “branch” for the selected root entity. Make sure it is visible in the window.
4. Click on the arrow tool to the right of the attribute drop down list. The attribute name should show in the Branch box.
5. Click on the Attach button to attach the attribute to the entity. If an error is made in attaching an attribute to the wrong entity or in the wrong order, use Detach to remove that attachment.
6. To create another attribute branch, click on the Clear fields button. Repeat steps 1-5. There is not a way to keep the entity in the Root box and just change the attribute in the Branch box. You have to clear it, then re-select the entity and select the attribute.

You must follow the order above. If you add an attribute first, it will put it into the Root box and you won’t be able to attach it to the entity. First entity, then attribute. Attach. Clear. Repeat for each attribute.

Be sure to attach attributes in the correct order relative to the data resource. If you do immediately discover a mistake, you can detach that branch and re-do it. While you can detach any branch you wish, you cannot surgically attach an attribute to an entity somewhere in the middle of the attribute list. It always appends to the list. This would require detaching backwards through the list to that attribute and re-attach forward in the correct order. Or you could ask a data manager nicely to fix it for you. Clicking on View Existing Branches will allow you to preview the attribute list for the selected entity and the ordering. It is a good practice to periodically inspect your entry using View Existing Branches as you go.

**Attaching a Domain to an Attribute**

Attaching a domain to an attribute is essentially the same process as outlined for attaching an attribute to an entity.

1. Clear the fields
2. Select the attribute
3. Click on the arrow next to the entity list. The attribute name should show in the Root box.
4. Click on the domain from the domain drop down list to select it.
5. Click on the arrow next to the domain list. The domain name should show in the Branch box.
6. Click on the Attach button.
7. Click on Clear fields to clear the selection and repeat for next domain.

As for attributes, enumerated domain values should be attached in the order you wish them to be, though there is no specific requirement for doing so.

As mentioned before, a domain can be attached to more than one attribute, Clicking on View Existing Branches will allow you to preview the domain list for the selected attribute and is a good way to verify all values are present.
Distribution

The Distribution tab is used to collect information about how the data resource is distributed to the public. Effectively, how the user obtains the data resource and pertinent information they might need to know. There can be more than one distribution in metadata, relative to multiple download sites and distribution methods. Keep in mind that the authoritative copy of the data must be distributed from a USGS public facing repository, with the default being the ASC, unless agreement has been made for use of other repositories. Data can also be distributed as a secondary download site through collaborator sites, at the discretion of the PI.

It is easiest to include the ASC distributor information in a template, then you don’t have to think about it other than documenting alternative distributors. The data manager will fill out portions of this section for you.

Distributors

This screen is used to define a listing of distributors for the data resource. You can select an existing distributor for edit or use Add to create one.

Figure 44: Distributors screen
**Distributor**

This screen is used to provide a citation for the distributor and define the parameters on how a user can obtain the data resource from that distributor. Distribution liability should use the appropriate USGS statement for the data resource. See Metadata Resources for the web link to standard statements.

![Distribution Information screen - ASC example](image)

- Click here to add/edit contact information for distributor.
- Standard liability statement(s) go here.
- Statement about the distributor’s role with the data set.
Contact Information

We prefer to use generic contact information for the ASC to prevent the likelihood of this information growing stale. Also, distribution questions may be IT related questions so the interceptor of an inquiry can redirect appropriately. Use the standard language for ASC contact information. Again, best to just use a template. For secondary distributors, this information will need to be provided. Note that email is considered optional, but should be provided if at all possible.

Figure 46: Contact Information screen for distributor
Ordering

Ordering is used to describe how the user can access the data resource from the distributor. You can have as many ordering processes for the distributor as you like, though generally there is only one.

**Standard Order Process**

You can describe the various digital forms the data can be accessed by. Your data manager can help with this section for data distributed through USGS channels. This screen is an example for an ASC standard zip package distribution. USGS cannot charge for distribution. Generally speaking, the intent is to describe the format the data is acquired in, not the format of the data itself.
**Digital Transfer Information**

This screen is used to describe what the user will get. Generally, the data is distributed in one or more zip packages. This example is for an ASC distribution and is best put in a template and edited as needed.

![Digital Form screen](image)

**Figure 49: Digital Form screen**

Optional but we like to provide this. Report in MBs.

The release DOI goes here. The DM will provide this.
Prerequisites & Availability

This section is optional but should be used to identify specialized SW that is needed in order to utilize the data. This is especially true for data that is commonly visualized.

![Prerequisites & Availability screen](image)

**Description of HW/SW and other requirements to utilize the data set.**

Optional. Not used for ASC distribution. May be relevant for other distributions.

**OK/Cancel**

This screen is only used to save the distribution information. **IMPORTANT:** Click OK here to save your distribution edits. Closing the Distribution Information window without clicking OK risks losing your edits. Note, this still does not save the edits to the file. You will still need to do a file save using the save button from the main screen.

![OK/Cancel screen](image)
**Metadata Ref**

This tab is used to provide contact information for the metadata record and is required. Again, we prefer to use a generic ASC citation which is best included in a template. The main information you will be concerned with is identifying whether you are using the core standard (CSDGM) or the Biological Profile.

![Figure 52: Metadata Ref](image)

- Set to CSDGM or Biological Profile.
- Set to 1999 for Biological Profile, 1998 for CSDGM.
- Date set by Metavist when file first saved. This should be edited to reflect last date of edit.
- Contact information for the metadata.
- Not used by ASC.

![Figure 53: Metadata Contact screen](image)
Extensions/Constraints & Security

The Extensions screen is not implemented and the Constraints and Security screen is not used since our metadata does not have security considerations.
Summary

Congratulations, you have made it through the metadata authoring process using Metavist. Hopefully this guide has helped get you up and running with using Metavist. Please see the Metadata Resources for other support resources in authoring metadata. If you have questions you can always call on a data manager, your data steward or anyone else with experience in writing metadata. If you find errors, omissions or better ways to do things, please drop me a note.