A Guide to the Data Elements

Most of the information you need to complete the Content Standard for Digital Geospatial Metadata (CSDGM) — also called the *Green Book* — is dealt with while you produce the data set. Consequently, your metadata is more accurate and much easier to create if you record the answers to questions when you actually produce the data. *Why deal with information twice?*

Remember, the purpose of creating metadata is to share your geospatial data with subsequent users. You shouldn't be using data unless you can answer the basic *who*, *what*, *when*, *where*, *why*, and *how* questions about it.

Keep these questions in mind while you produce your data set:

1 Identification Information

1. What is the title of the data set? (The title should include theme, place, time, purpose. Include any information on the series in which it is a part, or which edition it is.)

2. How are the data intended to be displayed?

3. What is included in the data set? (Write a brief summary.)
   - Why is the data set being created?
   - What are the objectives of the research of which producing this data set is a part?
   - How do you recommend that the data be used?
   - Are you concerned that the data may be misinterpreted? If so, list information that needs to be considered.)

4. What time period do the data represent?

5. When you publish your metadata, will you be finished with the data set? Or is it an ongoing work-in-progress?

6. What geographic area does the data set cover? Use decimal degrees to demarcate the area's west, east, north, and south boundaries. (This will create a rectangle. If you wish to mark out a more specific area, record the coordinates of your study area and see the *Green Book*, section 1.5.2, for more information.)

Don't Duck Metadata

❷ Data Quality Information

This section asks you to look a little closer at your data than you would probably want! Remember, as the creators of official geospatial data, your responsibility is great!)

1. How accurate is the identification of entities and assignments of values in your data set? How was accuracy assessed?
   - — Did you ground truth your findings?
   - — Did you compare your findings to another source?
   - — How accurate was the other source? If you have a statistic that measures the level of accuracy, include it with a description of how the statistic was obtained.

2. What geometric problems and drafting inconsistencies exist within the data set?
   - — Do the observations mean the same thing throughout the data set?
   - — What contextual factors should be considered by someone who wants to understand the data?
   - — Did the instrumentation or calibration change while observations were being made?
   - — Does it matter who made the observations? If so, report how many people made the observations.

3. Does the data set have topology? If so:
   - — What spatial relationships does the GIS recognize?
   - — Do lines intersect only where intended?
   - — Are any lines entered twice?
   - — Are all areas completely described?
   - — Are there any overshoots or undershoots?
   - — Are polygons intersected by the neat-line closed along the border?
   - — Are any polygons too small, or any lines too close?
   - — At what tolerance was cleaning, or "snap-to," function set?

4. Scale and legibility constraints require all data sets to generalize to some degree. What selection criteria, generalization, or definitions are used to determine what is included in the data set?
   - — Do you have minimum requirements for the classification of attributes?
   - — Did you omit outliers? (If you used, or manipulated, a standard geocode, such as the Federal Information Processing Standard, or FIPS, note it here.)

5. How closely does the data set's coordinate descriptions of features compare to their actual, "real-world" location?

   — What tests were used to determine the positional accuracy of your data set?

   — If you use a statistic to measure positional accuracy, include a description of how the statistic was derived.

   — If your data includes vertical positions (elevation, depth), record an assessment of vertical positional accuracy. (Consider such things as the accuracy of data collection methods, conversion algorithms used, etc.)

6. Did you use other data sources to create your data set? If so:

   — Where did the data come from? (Include the citation information for any data sets that were used to create the one you're currently working on.

      - If other data sources were unpublished, where did you get them?
      - At what scale were your source data collected?
      - Is there a discrepancy in the time period of content? If so, include a description of how you compensated for the discrepancy.

   — How were the data modified?

      - Did you have to compensate for data that were collected at different scales?
      - Did you manipulate the map projection, or grid system, of your source data?
      - When did your process step occur?
      - Did a third party perform the processing step? If so, include contact information of the third party.

If your process step generated any data sources that stand on their own, it is a good idea to create metadata for them, too.

After recording this information, it might be a good idea to revisit question #3 in Section 1, *Identification Information*: "Are you concerned that the data may be misinterpreted? If so, list information that needs to be considered."

❸ Spatial Data Organization Information

1. If you were at a cocktail party, how would you describe the location of your study to someone? (You probably would not use decimal degree bounding coordinates!) For example: *Northwest North Carolina*, or *Watauga County*.

2. What system of objects is used to represent space in the data set? (point, vector, raster)

3. If you would like to include information on the specific spatial objects used to locate point, line, and area locations in your data set, record them here. (See the *Green Book*, section 3.3.1.1-3.4.4 for more information about recording point and vector object types, vector product format, or raster grid cell descriptions.)

❹ Spatial Reference Information

1. What kind of coordinate system is used to locate entities in your data set?
   - — **Geographic** = degrees/min/sec or decimal degrees
   - — **Planar** = map projections, grid coordinate systems, or local planar coordinate systems
   - — **Local** = any coordinate system with a point of reference that is not aligned with the earth

2. **If your coordinate system is *geographic*** *(latitude/longitude)*:
   - — What are the units of measure used for your data set's longitude/latitude coordinates (Decimal degrees, decimal minutes, decimal seconds, etc.)?
   - — What is the minimum difference between two adjacent points? (This is another question of scale; for example, if two entities lie within one minute of each other in the "real world," is your data set's scale such that this distance is recognized?)

3. **If your coordinate system is *planar*** (map projection, grid coordinate system, local coordinate system):
   - — What is the name of the coordinate system used?
   - — How does the planar system you used represent horizontal positions (for example: coordinate pair, distance and bearing, row and column)?
   - — What are the units of measure used for distances? (During the process of creating your data set, if you record the name of the system, its method of horizontal positioning, and units of measure, that is enough. However, you will need to see the *Green Book* for more information about how to properly document planar data according to the CSDGM.)

4. **If your coordinate system is *local:*** (any coordinate system that is not aligned with the earth)

   — How does the local coordinate system locate a point spatially? (Describe how the coordinate system works)

   — How can the system be georeferenced? If applicable (and it probably is), answer questions 5 and 6.

5. What **horizontal datum** is used by the coordinate system?

6. What **ellipsoid** is used by the coordinate system? (If your datum is NAD 27, your ellipsoid is "Clarke 1866." If your datum is NAD 83, your ellipsoid is "Geodetic Reference System 80." See the *Green Book*, section 4.1.4.2, for further explanation.)

   — What is the radius of the equatorial axis of the ellipsoid? (semi-major axis)

   — What is the ratio of the difference between the equatorial and polar radii of the selected ellipsoid? (flattening ratio) (semi-major axis radii, and flattening ratios, of several major ellipsoids, are listed in section 4.1.4.4, of the *Green Book*)

(If your data set uses a vertical coordinate system for representing elevation or depth, you will be required to answer questions 7 through 9.)

7. What elevation, or depth, datum is assumed by the coordinate system? (This communicates the point of reference for altitude or depth measurement.)

8. Do scale constrictions limit the resolution of your altitudinal, or depth, data? What is the minimum elevation, or depth, difference between two points?

9. Were elevations, or depths, interpolated from sample points? Or, does each horizontal coordinate pair (x, y) have a vertical (z) coordinate?

❺ Entity and Attribute Information

Keep a record of all cartographic decisions!

We all have dealt with these concepts when creating data sets. However, due to terms and definitions used by the FGDC, a quick explanation is needed.

— An **entity** is an object represented in your data set.

— An **attribute** is a defined characteristic of that entity.

— **Attribute values** are the domain of possible classifications into which specific entities are assigned. Attribute values can be nominal, ordinal, unclassifiable (addresses), or can use an established standard (such as my examples below, which use the Cowardin wetland and deepwater classification system).

---

1. What entities are included in your data set? Include a list of how each entity is represented in your data set and a definition of each entity.

   — What is the source for your entity definition? For example:

   - **entity** = wetland
   - **entity definition** = "Wetlands are lands transitional between terrestrial and aquatic systems where the water table is usually at or near the surface or the land is covered by shallow water."
   - **entity definition source** = Cowardin, L.M., V. Carter, F. Golet, and E. LaRoe. 1979. *Classification of wetlands and deep water habitats of the US*. U.S. Fish Wildlife Service. pp. 103.

2. What is the name of each attribute in your data set?

   — What does each attribute explain?

   — What is the source of your attribute definition? For example:

   - **attribute label** = wetlands classification
   - **attribute definition** = Wetlands are classified by their proximity to the coast, and hydrologic conditions
   - **attribute definition source** = Cowardin, L.M., V. Carter, F. Golet, and E. LaRoe. 1979. *Classification of wetlands and deep water habitats of the US*. U.S. Fish Wildlife Service. pp. 103.

3. Include a list of attribute values, and a description of each domain. (Basically, all you have to do is transfer it from the scrap piece of paper on your desk to this document!)

   — Describe how entity attributes were classified.

   — If you used an established standard for domain values, include the name of the standard and the definition of each attribute value.

   - **Marine** = open ocean and associated coastline
   - **Estuarine** = salt marshes and brackish tital water
   - **Riverine** = rivers, creeks, and streams
   - **Lacustrine** = lakes and deep ponds
   - **Palustrine** = shallow ponds, marshes, and swamps.

4. Is the domain a list of nominal values, ordinal values, defined by a set of codes, or unrepresentable?
   - — If **nominal**, include a definition of each value and the source of the definition.
   - — If **ordinal**, include the minimum and maximum classification values.
   - — If you used a **set of codes**, include the code set name and code set source.
   - — If **unrepresentable**, define what the attribute value is. (The elements required by the *Green Book* vary by what level of measurement is indicated by the domain values. See section 5.1.2.4 for an explanation.)

5. What units of measure are used for the classification of each entity?

❻ Distribution Information

Few fields of this section will ever change. Therefore metadata creators usually need only complete it once and use cached copies for subsequent metadata entries.

1. Is the contact a person or an organization?

2. What is the mailing address, telephone number, and e-mail?

3. Is there a label or a catalog number by which the contact can better identify and retrieve this data set?

4. Include a statement of liability assumed by the distributor.

5. What is the standard ordering procedure? (See the *Green Book*, section 6.4, for required fields for digital ordering.)

6. Is there a fee for obtaining the data set?

❼ Metadata Reference Information

Section 7 should easily be filled out on the day you write your metadata!

Modified from Peter Schweitzer's *Metadata in Plain Language*, http://geology.usgs.gov/tools/metadata/tools/doc/ctc/. *Many* thanks to Peter for his *many* contributions! Many questions are taken directly from the FGDC's CSDGM, version 2.