From the F[i]GDC and the National Metadata Cadre'…

Ten Most Common Metadata Errors

Ten Most Common Metadata Errors

---

10. Defining your data set too finely or too broadly

It's easy to become overwhelmed trying to individually document every data table and resource. On the other hand, trying to cover all of your data resources with a single metadata record will drive both you and your data users crazy. A good rule of thumb is to consider how the data resource is used – as a component of a broader data set or as a stand-alone product that may be mixed and matched with a range of other data resources.

9. Using incorrect State Plane Coordinate System Zone Identifier values

The default SPCS Zone Identifier (4.1.2.2.4.1) values for some software products are based upon early BLM values rather than the FIPS Code prescribed by the CSDGM.

8. Confusing 'Currentness Reference' with 'Publication Date'

While the Currentness Reference (1.3.1) may refer to a publication date it is actually a qualifier to Time Period of Content (1.3). Does the time period refer to the date/time of data capture or *ground condition* as in photography or field data collection? Does it refer to the date the information was officially *recorded* as in a deed? Does it refer to a *publication* date as in a '1978 USGS Topo map'? Basically, the idea is to let prospective users know how well you are able to 'nail' the actual time period of content.

7. Misunderstanding resolution

Who could blame us? The purpose of these fields is to indicate how coarsely or finely information was recorded. For example

- Latitude Resolution (4.1.1.1) and Longitude Resolution (4.1.1.2)
  These values represent the minimum possible difference between coordinate values. For example:

	resolution (4.1.1.1 or 2)	geographic coordinate units (4.1.1.3)
30° 30' 30"	0.0028 (1° / 360")	degrees, minutes, seconds
30° 30' 30.01"	0.000028 (1° / 36,000")	degrees, minutes, decimal seconds
30.00001°	0.00001 (1° / 100,000)	decimal degrees

- Abscissa/Ordinate Resolution (4.1.2.4.2.1 and 2)
  These values represent the minimum difference between X (abscissa) and Y (ordinate) values in the planar data set. For raster data, the values normally equal pixel size, e.g. 30 (TM). For vector data, the values usually indicate the 'fuzzy tolerance' or 'clustering' setting that establishes the minimum distance at which two points will NOT be automatically converged by the data collection device (digitizer, GPS, etc.). NOTE: units of measures are provided under element Planar Distance Units (4.1.2.4.4) and would be 'meters' for the TM example provided and likely millimeters for the vector example.

6. Putting too much faith in metadata tools

Human review is the only thing that matters. The tools are there to help, remember: 'garbage in - garbage out'.

5. Taking the minimalist approach

A common overreaction to the expansive nature of the CSDGM is to adopt 'minimal compliance' as an operational approach. Limiting your documentation to the 'required' portions of Sections 1 and 7, or even all 'required' fields, will limit the value of your effort and the metadata records you produce. Instead, identify those fields that apply to your organization and data, and create functional templates, or subsets, of the CSDGM.

4. Understanding assessments of consistency, accuracy, completeness, and precision

<u>Section 2. Data Quality Information</u> is intended to provide a general assessment of the quality of the data set. This represents the 'Achilles heel' for many RS/GIS professionals. Consider it an 'opportunity' to get to know your data set. A brief summary:

- <u>Attribute Accuracy Report</u> (2.1.1)
  Assessments as to how 'true' the attribute values may be. This may refer to field checks, cross-referencing, statistical analyses, parallel independent measures, etc. Note: this does NOT refer to the positional accuracy of the value (see 2.4).
- <u>Logical Consistency Report</u> (2.2)
  Assessments relative to the fidelity of the line work, attributes and/or relationships. This would include topological checks, arc/node structures that do not easily translate, and database QA/QC routines such as: Are the values in column X always between '0' and '100'? Are only text values provided in column Y? For any given record, does the value in column R equal the difference between the values provided in columns R and S?
- <u>Completeness Report</u> (2.3)
  Identification of data omitted from the data set that might normally be expected, as well as the reason for the exclusion. This may include geographic exclusions, 'data was not available for Smith County'; categorical exclusions, 'municipalities with populations under 2,500 were not included in the study'; and definitions used '*floating marsh* was mapped as *land'.*
- <u>Positional Accuracy</u> (2.4)
  Assessments of horizontal and/or vertical positional (coordinate) values. Commonly includes information about digitizing (RMS error), surveying techniques, GPS triangulations, image processing or photogrammetric methods.
- 'Precision'
  An indication as to how 'finely' your data was recorded, such as digitizing using single or double precision. Note that the precision of the value in no way reflects its accuracy or truthfulness.

3. Glossing over Section 5. Entity and Attributes

Another of the GIS professional's 'Achilles tendons', this section maps out data content and should be a product of your data design effort.

- Use the relational database format as a *guide*:
  - <u>Entity Label</u> (5.1.1.1) – Table Title
  - <u>Attribute Label</u> (5.1.2.1) – Column Titles
  - <u>Attribute Domain Values</u> (5.1.2.4.X) – Recorded values within each column

- Domain Types – set of possible data values of an attribute
  Enumerated Domain (5.1.2.4.1)
    A defined pick list of values
    Typically categorical such as road types, departments, tree types, etc.
  Range Domain (5.1.2.4.2)
    A continuum of values with a fixed minimum and maximum value
    Typically a numeric measure or count, may be alphabetic (A–ZZZ)
  Codeset Domain (5.1.2.4.3)
    A defined set of representational values
    Coding schemes such as FIPS County Codes, or Course No. (GEOG 1101)
  Unrepresentable Domain (5.1.2.4.4)
    An undefined list of values or values that cannot be prescribed
    Typically text fields such as individual and place names
- Entity Attribute Overview (5.2.1)
  A summary overview of the entities/attributes *as* outlined in *either* Detailed Description (5.1) *or* an existing detailed description cited in Entity Attribute Detail Citation (5.2.2). Note that the field should not be used as a stand-alone general description.

2. Thinking of metadata as something you do at the end of the data development process

Metadata should be recorded throughout the life of a data set, from planning (entities and attributes), to digitizing (abscissa/ordinate resolution), to analysis (processing history), through publication (publication date). Organizations are encouraged to develop operational procedures that 1) institutionalize metadata production and maintenance, and 2) make metadata a key component of their data development and management process.

1. Not doing it!

If you think the cost of metadata production is too high – you haven't compiled the costs of **not** creating metadata: loss of information with staff changes, data redundancy, data conflicts, liability, misapplications, and decisions based upon poorly documented data.

---

[i]This is a product of the Federal Geographic Data Committee (FGDC) Metadata Education Program and the National Metadata Cadre'. Much of the information was collected during metadata workshops, meetings and presentations and represents the comments and contributions of many individuals. Specific contributors include: Lynda Wayne, editor, FGDC contractor • Anne Ball, NOAA, Coastal Services Center • John Bocchino, New Jersey Department of Environmental Protection • Chris Cialek, Minnesota State Planning • George Lienkaemper, USGS, Forest and Rangeland Ecosystem Science Center • Jackie Olson, USGS, Water Resources Division • Richard Pearsall, FGDC • Hugh Phillips, 3001, Inc. • Nina Savar, Northeastern Illinois Planning Commission • Peter Schweitzer, USGS, Geologic Division • Sheryl K. Soborowski, Virginia Tech • Bruce Westcott, RTSe, Inc.