Documenting Tabular Geospatial Data
A Metadata Quick Guide

Background:
The term 'geospatial data' is most commonly used to reference maps, photographs, satellite images and Geographic Information System (GIS) products. However, vast amounts of geospatial data are collected in common tables and databases. This tabular data may include a specific column for geographic coordinates but it is far more common that geospatial location is referenced by street address, administrative boundaries such as school districts, geophysical features such as watersheds and other ''mappable' attributes. At some level, most tabular data is geospatial. Though the geospatial character of the dataset features may reference an area rather than a specific point, the data can, none the less, be tied to a location on the earth.

Geospatial Metadata:
Digital data resources are documented by a set of descriptive elements referred to as metadata. *Geospatial* metadata includes special elements needed to effectively describe the geospatial nature of digital data. Metadata can be considered a library record for digital data. It captures basic information such as the data set title, author, abstract, time period of content and geographic extent. It also records information about how the dataset was developed and how it is maintained. Metadata supports data and project management by enabling both data managers and data consumers to:
- identify and locate needed data resources
- assess the applicability and quality of an available data resource
- manage and maintain geospatial data
- archive data for historical analyses and preserve data investments.

Content Standard for Digital Geospatial Metadata
The *Content Standard for Digital Geospatial Metadata* (CSDGM) is the current US federal standard used to ensure that geospatial data are documented in a consistent manner. There are efforts underway, through the International Standardization Organization (ISO), to adapt the CSDGM to an international metadata standard but in the interim, the CSDGM remains the most common standard within the US.

The CSDGM was designed to apply to all forms of digital geospatial data including:
- GIS data
- imagery
- geospatial tabular data (tables and databases)
- digital cartographic products.

The CSDGM includes many elements specific to each of these data types and is, as a result, an extremely 'robust' standard comprised of more than 900 elements. However, not all elements will apply to each data type and it is the responsibility of the geospatial data manager to identify those elements pertinent to the data format, data themes and organizational objectives. Once the pertinent elements are identified, the geospatial manager can build templates that 'carve down' the standard and present only pertinent elements and, where possible, provide picklists or guidance on populating the element.

Methods and Tools for Creating Tabular Geospatial Metadata
Metadata should be created at the onset of a data development project and maintained through the data life cycle. By making metadata part of the data development process, the data manager is able to:
- outline the data criteria, sources and processing
- effectively capture critical information about the data
- monitor data set development.

Specific methods for integrating metadata into data development are outlined in the FGDC guidance document *Institutionalize Metadata Before It Institutionalizes You* available from the FGDC Metadata Publications website at:
http://www.fgdc.gov/metadata/metadata-publications-list

Several metadata editors are available to aid in the creation of geospatial metadata. While most GIS and image processing software applications include editors capable of automatically capturing data set properties (projection, attribute labels, etc.) and forms to record descriptive information (purpose, attribute definitions, etc.), spreadsheet and database applications seldom provide such tools. However, several commercial and freeware editors are available to support data documentation outside of the data development application. The USGS-sponsored metadata editors TKME (for the Windows operating environment) and XTME (for the UNIX operating environment) are well-maintained applications that balance portability with usability. The editors are available free of charge from the USGS Geology Division:
http://geology.usgs.gov/tools/metadata/

Additional information about metadata editor and validation software applications is available from the FGDC Metadata Program:
http://www.fgdc.gov/metadata/geospatial-metadata-tools

Applying the CSDGM to Tabular Data
The FGDC provides specific guidance on the metadata elements considered essential to effective data documentation. These elements are outlined in the *Essential Geospatial Metadata* document maintained at the FGDC Metadata Publications website (http://www.fgdc.gov/metadata/metadata-publications-list). Data managers are encouraged to use this document as a starting place for the creation of custom metadata templates. In the case of tabular data, the metadata template must go beyond essential elements and focus should be placed on the detailed documentation of the dataset 'entities and attributes'.

Documenting Entities and Attributes
For *most* tabular data sets, the data table is the 'entity' and the column headings are the 'attributes', and the data values displayed within the row are the 'attribute values'. In more robust or unique data table configurations, such as object oriented databases, the attributes can be defined as the 'features' and the attribute values would be the 'instances' of features.

Within the CSDGM, entities and attributes can be described in two different ways.

1. 'Detailed Description'
 The Detailed Description is used if the dataset has not been documented elsewhere. Consider this your 'opportunity' to document your entity and attribute labels and values so that others can apply your data in an appropriate manner.

   The elements of the 'Overview Description' include:

   Entities
   Tabular data entities are most commonly the tables used organize the information and are documented using the following key CSDGM elements:
   Entity Type Label
   ▪ name of the table
   ▪ EXAMPLE: 'Paduka County Assets (Roads) Management
   Entity Type Definition
   ▪ description of the table
   ▪ EXAMPLE: State, 'County and Local roads name, pavement type, dimensions, direction, number of lanes, start and end points.'
   Entity Type Definition Source:
   ▪ formal or informal source of definition
   ▪ EXAMPLE:
     o informal = the organization writing the metadata record
     o formal, if the definition is taken from some published document: the organization that published the document, e.g. 'State Highway Administration'

2. 'Overview Description'
The Overview Description is used only if the entities and attributes are well documented by some other publication or website. Common examples include the use of nationally published data tables such as those included in the Census Summary Files, National Hydrography Dataset (NHD), National Health Interview Survey (NHIS) and National Archive of Criminal Justice Data.

The elements of the 'Overview Description' include:
Entity and Attribute Overview
- a brief description of the content and/or a listing of the entity and attribute labels
- EXAMPLE: 'A detailed description of each interstate and state road including name, pavement type, dimensions, direction, number of lanes, start and end points.'

Entity and Attribute Detail Citation
- unlike most CSDGM 'Citations' this is an informal reference to the publisher and should include the name of the publisher and an available URL to access the full documentation
- EXAMPLE: US Geological Survey, http://edc.usgs.gov/products/map/dlg.html

Attributes
Tabular data attributes are most commonly the column headings used to organize the information and are documented using the following key CSDGM elements:
Attribute Label
- column heading
- EXAMPLE: 'Road Condition '

Attribute Definition
- description of information documented in the column

EXAMPLE: 'The conditions of the road surface in relation to pavement preservation*:*
- informal = the organization writing the metadata record
- formal, if the definition is taken from some published document: the organization that published the document, e.g. ' Definitions for pavement preservation from US Department of Transportation, Federal Highway Administration: Pavement Preservation Definitions , http://www.fhwa.dot.gov/pavement/preservation/091205.cfm

Attribute Domain Values
For each attribute, a specific type of information, or domain of values, can be entered. Some domains are completely open, such as 'Name' while others may be limited by a picklist or max/min.

The following domain types are used to document attribute domain values:

Attribute Domain Types:
  Enumerated Domain
    ▪ description: a predefined list of valid values
    ▪ EXAMPLE: 'unpaved, gravel, shell, asphalt, concrete'
  Range Domain
    ▪ description: a maximum and minimum value for asphalt depth.
    ▪ EXAMPLE: 0-100 percent
  Codeset Domain
    ▪ description: any published set of values that are defined elsewhere
    ▪ EXAMPLE: alligator, depressions, patching, etc
  Unrepresentable Domain
    ▪ description: any domain set for which the values simply cannot be predeterimined or defined
    ▪ EXAMPLE: Road names in Paduka County

The following CSDGM elements are used to document each of the domain types:
  Enumerated Domain
    Enumerated Domain Value (repeated for each valid value)
      ▪ a single valid value
      ▪ EXAMPLE: 'unpaved'
    Enumerated Doman Value Definition
      ▪ description of the single valid value
      ▪ EXAMPLE: 'bare ground'
  Range Domain
    Range Domain Minimum
      ▪ the minimum valid value
      ▪ EXAMPLE: '0'
    Range Domain Maximum
      ▪ the maximum valid value
      ▪ EXAMPLE: '100'

  Codeset Domain
    Codeset Name
      ▪ published name of the codeset
      ▪ EXAMPLE: 'Paduka County Paving Conditions Definitions'
    Codeset Source
      ▪ name of the organization that publishes the codeset and, if possible, the URL of the online source for the codeset
      ▪ EXAMPLE: 'Paduka County http://paduka_ky.gov/PavingConditionsDefinitions.pdf

  Unrepresentable Domain
    Unrepresentable Domain
      ▪ description of the values and reasons why they cannot be represented

- EXAMPLE: 'The names of the roads do not exist in a known, predefined set'

Learn More About Geospatial Metadata

The FGDC Website (http://www.fgdc.gov/metadata) offers additional information about geospatial metadata.

- Value and Making the Business Case
- Standards, Profiles and Extensions
- Software Applications
- Guidance Documents, Brochures and Other Publications
- Training Opportunities and Instructors
- Working Group
- Grants
- Links to NSDI Stakeholder Resources
- Links to the USGS Geology Metadata FAQ.

The FGDC NSDI Online Training Initiative (http://www.fgdc.gov/training/nsdi-training-program/online-lessons) provides the following metadata presentations for download or online viewing:

- What is Metadata?
- Value of Metadata
- North American Profile Development for ISO 19115 Geospatial Metadata

Additional metadata training materials developed by NSDI Stakeholders are available from:
http://www.fgdc.gov/training/training-materials#stakeholderstraining

For individual assistance, contact the FGDC Metadata Coordinator at:
    metadata@fgdc.gov