



FGDC Don't Duck Metadata

Metadata Quick Guide

A short reference guide for writing quality metadata

- Formerly published as "Metadata Cliff Notes", the document has been updated to include more specific information regarding 'Online_Linkage', 'Progress', and 'Theme_Keyword' elements.
- **March 1, 2005** Corrected ISO Topic Category example 'soil fertility' (page 9) and revised capitalization of ISO Topic Category Names (pages 9-10).

General Recommendations

Use the 'Green Book' (CSDGM Workbook)

This resource is filled with many rich explanations and examples

For DOQQ's and other large data sets with many tiles...

Create a single metadata record for the entire data set if time and resources are short, BUT try to create new, more accurate metadata as you build regional subsets. Eventually, try to create independent files for each image.

Section 1: Identification

Citation (1.1)

Originator (8.1)

It is recommended that you indicate the party responsible for the data set. While that is most commonly the organization that developed the data set, in some cases, it is not. For example, if a county planning department hires a contractor to build a street centerline road file, the planning department, not the contractor should be identified as the Originator. The contractor should be fully cited using the Data_Set_Credit (1.11) element, e.g. *'this data set was developed for the Wayne County Planning Department by Smith Engineering, Inc'*.

Publication_Date (8.2)

The date that the data was published or otherwise made available. Remember format: YYYY/MM/DD.

Title (8.4)

- Minimum - where, what, when,
- Best practice - who, why, resolution, filename, source
e.g. *Aquifer Systems and Recharge Potential in Louisiana from LDEQ source data, Geographic NAD83, LOSCO (1999) [aqrgeog3dpdeq]*

Online_Linkage (8.10)

As 'repeatable' elements, **Online_Linkage** (Citation Information) and **Network_Resource_Address** (6.4.2.2.1.1.1) are used to provide access to a variety of data download, data clearinghouse, and web-mapping services. Use this field to fully represent your geospatial data access and distribution capabilities by providing complete URLs and necessary information to indicate the nature of the weblink using the following style guidance:

- OGC Web Map Service (WMS) links include a 'getmap' request with a layer name, version, preferred image format, and preferred SRS, at a minimum:
<http://server/service?REQUEST=getmap&VERSION=1.1.0&LAYERS=roads&FORMAT=image/gif&SRS=EPSG:4326>

- ArcIMS “Image” services using a URL-like request. If you pasted this request in a browser you will not connect to an ArcIMS server since it does not permit this style of request, however it contains enough information to allow geodata.gov to connect to an ArcIMS service: http://<server>/image/<service_name> will be assigned as Live Map, ArcXML Image service, where server URL is <server> and service name is ArcIMS <service_name>. The sub-path “/image/” must be present in the URL
- direct download sites include URLs, that start with either ftp:// or http:// and point to filenames with .zip, .tgz, .gz, .dxf, or .e00 extensions.

Abstract (1.2.1)

Be sure to include

- general content and features
- data set form (GIS, CAD, image, Dbase)
- geographic coverage (county/city name)
- time period of content (begin and end date or single date)
- special data characteristics or limitations

Supplemental_Information (1.2.3)

A comment field in which you can:

- place information that is not elsewhere covered
- ‘front’ important information such as related studies, data set limitations, and notifications

Time_Period_of_Content (1.3)

The relevant date of the data content. Can be a single date, multiple dates, or a range of dates.

Currentness_Reference (1.3.1)

The context for the Time_Period_of_Content. For example: an orthophotograph may have been compiled and delivered in June (publication date) but flown in February (ground condition).

Progress (1.4.1)

The status of the data set, this field has a fixed domain of: “Complete”, “In Work”, and “Planned”.

Note that federal agencies must create metadata for planned data acquisitions estimated at a cost of \$500,000 or greater (as of FY05) to enable discovery by potential data development partners.

Theme_Keyword (1.6.1.1)

Include broad and specific terms and use controlled vocabularies (thesauri) when possible.

- Include at least one ISO Topic Category (see page 8) referencing the associated *Theme_Keyword_Thesaurus* as ‘ISO 19115 Topic Category’
- Include additional descriptive terms to qualify Topic Category

Place_Keyword (1.6.2.1)

Include specific and regional references such as:

- city or county name
- state
- state acronym
- regional descriptions and references e.g., Appalachia, Puget Sound, DelMar Peninsula, etc.

Stratum_Keyword (1.6.3)

For use in atmospheric, geologic, and oceanographic data, e.g., ionosphere, surface, seafloor

Temporal_Keyword (1.6.4)

For use in scientific and historical data, e.g., diurnal, Ming dynasty, Machine Age

Access_Constraints (1.7)

Any restrictions or legal prerequisites to accessing the actual data set. Commonly applies to data sets that are exempt from public records laws such as endangered species, personal health, and intellectual properties.

Use_Constraints (1.8)

Any restrictions or legal prerequisites to using the data set. Common constraints include:

- must read and fully comprehend the metadata prior to data use
- acknowledgement of the Originator when using the data set as a source
- sharing of data products developed using the source data set with the Originator
- data should not be used beyond the limits of the source scale
- the data set is NOT a survey document and should not be utilized as such

Point_of_Contact (1.9)

The individual or organization that is knowledgeable about the data set and should be contacted with questions.

Data_Set_Credit (1.11)

Identify others that should be recognized for their contributions to the data set. This includes data development contractors as discussed, above, for Originator.

Native_Data_Set_Environment (1.13)

Optional but highly recommended

- software and version
- operating system and version
- platform

Section 2: Data Quality

Attribute_Accuracy_Report (2.1.1)

How sure are you that it IS a pine tree?

Assessments as to how 'true' the attribute values may be. May refer to field checks, cross-checks with other documents, statistical analysis of values, and parallel independent measures. It does NOT refer to the positional accuracy of the feature.

Logical_Consistency_Report (2.2)

Did you check for bad values and conditions?

Tests used to check for data inconsistencies including topological checks (clean and build), and data base QA/QC routines such as: Are the X values always between '0' and '100'? Are all 'Y' values text format? Does value Z always equal the sum of values 'R' and 'S'?

Completeness_Report (2.3)

Is there anything I might expect to be in the data set that isn't?

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 the South Shores neighborhood'; categorical exclusions 'municipalities with populations under 1,000 were not included'; and definitions used 'floating marsh was mapped as land'.

Positional_Accuracy_Report (2.4)

How sure are you that the pine tree is where you say it is?

Assessments as to the horizontal and/or vertical location of the feature. May refer to field checks, Maximum Allowable PDOP, survey quality, cross-checks with other locational references, etc.

Process_Step (2.5.2)

Can be a single collective description or individual process steps based upon;

- stages of processing
- incorporation of sources
- project milestone

Process_Contact (2.5.2.6)

The individual responsible for the data processing and 'putting' the data together.

Cloud_Cover (2.6)

- leave blank for GIS and digital map files
- include values for imagery and photography

NOTE: this fields requires an integer, text responses should not be used.

- '0' through '99' indicate percent of the image obscured by cloud cover

- '100' indicates the value is unknown.

Section 3: Spatial Data Organization

Indirect_Spatial_Reference (3.1)

Any precise method of locating the data sans coordinates. Includes:

- Geographic Names Index System (GNIS) place names
- Public Land Survey System (PLSS) locations
- Federal Information Processing System (FIPS) location codes

Direct_Spatial_Reference_Method (3.2)

Indicate 'vector' or 'point' or 'raster'. Cannot select more than one.

SDTS_Point_and_Vector_Object_Type (3.3.1.1)

- for GIS files use 'Autocapture' feature of SMMS or ArcCatalog to populate
- see SDTS Definition Object Types at http://mcmcweb.er.usgs.gov/sdts/SDTS_standard_nov97/part1b10.html#152231

Section 4: Spatial Reference

Horizontal_Coordinate_System_Definition: (4.1)

Description of the reference frame for horizontal position. Can be **one** of the following:

- geographic
Latitude / Longitude
- map projected
Albers, Lambert, Transverse Mercator, Polar Stereographic, etc.
- grid coordinate system
UTM, State Plane, ARC, Universal Polar Stereographic, etc.
- local planar
Coordinate Geometry (COGO), rows/columns, etc.
- locally-defined
oblique photography, unrectified imagery

Abscissa_Resolution (4.1.2.4.2.1) / Ordinate Resolution (4.1.2.4.2.2)

The smallest distance that can exist between two points. The value is almost always the same for both the X axis (abscissa) and the Y axis (ordinate) but may differ for non-square pixels.

- Vector data
This is commonly 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.)
- Raster data
The values normally represent the pixel size, e.g. for Thematic Mapper (TM) imagery, the value would be '30'.

Note: this must be a real number and the units of measure are recorded as Planar_Distance_Units (4.1.2.4.4) (see next item).

Planar_Distance_Units (4.1.2.4.4)

The units of measures for the Coordinate_Representation (abscissa/ordinate resolution) or the Distance_and_Bearing_Representation. For the TM example provided above the units of measure would be 'meters. For the fuzzy tolerance example provided above, the units of measure would commonly be 'millimeters'.

Section 5: Entity and Attributes

Detailed_Description (5.1)

Provide a detailed description of your database if not documented in another form such as a data dictionary or data specification manual.

Attribute_Domain_Values (5.1.2.4)

Domain types

- Enumerated Domain
a defined set of possible values, a picklist
example: Anderson land cover classes
- Range Domain
a sequence, series, or scale that has defined maximum and minimum values -
can be numeric or alphabetical
example: date fields
- Codeset Domain
any published codeset
examples: USGS Digital Line Graph codes, FIPS codes
- Unrepresentable Domain
any value that is not prescribed
example: names

Overview_Description (5.2)

Provide an overview description if:

- your database is well-documented as a data dictionary, data specification manual, or some other format, AND you can provide data consumers a citation for the document and, if applicable, a website link to the document.
- your database is minimal and you can adequately describe in a short descriptive paragraph. For example, for a black and white orthophotograph, you may want to indicate that each pixel will have a gray scale value between 0 (black) and 255 (white). Be sure to explain any unclear attribute labels and codes.

Section 6: Distribution Information

Distributor_Contact (6.1)

The individual or organization that distributes the data.

Distribution_Liability (6.3)

A statement of the liability assumed by the Distributor. A legal-like section that may:

- deny liability if the data are incorrect, incomplete, or misused
- limit third party distribution of the data set

Section 7: Metadata Reference

Metadata_Date (7.1)

The date that the metadata is written or completed. Like other date fields, it can be a single date, multiple dates, or a range of dates.

Metadata_Contact (7.4)

The individual or organization that is responsible for the metadata for the data set.

Metadata_Standard_Name (7.5)

Content Standard for Digital Geospatial Metadata

Metadata_Standard_Version (7.6)

As of Oct 2002: *FGDC-STD-001-1998*

Metadata_Access_Constraints (7.8)

Restrictions and legal prerequisites for accessing the metadata (not the data). With the exception of classified information and intellectual properties, the response is almost always 'none'. Even if a data set is exempted from public record laws (endangered species locations, personal health data, etc.) the metadata is typically fully accessible.

Metadata_Use_Constraints (7.9)

Restrictions and legal prerequisites for using the metadata (not the data) after access is granted. This may applicable for the protection of privacy or intellectual properties. Note that though a data set may be exempt from public access, the metadata seldom contains any protected information such as the location of an endangered species nesting site or the address of an AIDS patient.

ISO 19115 Topic Categories from ISO/DIS 19115

Preparing for the international metadata standard:

Theme Keywords and the ISO Topic Categories

The International Organization for Standards (ISO) metadata standard (ISO 19115) provides a set of Core metadata elements that must occur in every national profile/implementation. Most of these elements either map to existing CSDGM metadata elements or represent properties of the data that can be determined and populated using a data integrated metadata tool. *Topic Category* is the only mandatory element of the ISO core metadata set that requires new information that cannot be directly captured from the data. The following 19 subject headings represent the domain for the *Topic Category* element.

If your metadata creation software provides a pick list of Topic Category related terms

simply select the pick list terms that apply and the software will insert the related Topic Category Name and/or Code. If creating data using the **Geodata.gov** metadata publisher, you will be asked to select a *Primary Theme*. The *Primary Theme* options are based upon the *ISO Topic Categories* below but the names have been altered to provide greater context, e.g., Geodata.gov Primary Theme 'Cultural, Society, and Demographic' will be captured in the Theme_Keyword metadata element as ISO Topic Category Name 'Society'.

If your metadata creation software does not provide a list of subject headings based upon the ISO 19115 Topic Category, include the *Topic Category Names* (as presented below) as *Theme_Keywords* and cite your related *Theme_Keyword_Thesaurus* as: 'ISO 19115 Topic Category'. The FGDC intends to develop CSDGM to ISO translation software that will insert the *Topic Category Code* when the *Topic Category Name* is found, however, those wishing to include the *Topic Category Code* as a *Theme_Keyword* can do so using the same *Theme_Keyword_Thesaurus*: 'ISO 19115 Topic Category'.

Include **all** pertinent Topic Category Names, e.g.,:

business districts = *boundaries* and *economy*
toxic release inventory = *environment* and *health*
soil fertility = *geoscientificInformation* and *farming*

ISO Topic Category Name, ISO Topic Category Code

farming, 001

rearing of animals and/or cultivation of plants
e.g., agriculture, crops, livestock

biota, 002

flora and/or fauna in natural environments
e.g., flora and fauna, ecology, wetlands, habitat

boundaries, 003

legal land descriptions
e.g., political and administrative boundaries

climatologyMeteorologyAtmosphere, 004

processes and phenomena of the atmosphere
e.g., processes and phenomena of the atmosphere

economy, 005

economic activities, conditions, and employment
e.g., business and economics

elevation, 006

height above or below the earth's surface

e.g., altitude, bathymetry, dem's, slope, derived products

environment, 007

environmental resources, protection, and conservation

e.g., natural resources, pollution, impact assessment, monitoring, land analysis

geoscientificInformation, 008

information pertaining to the earth sciences

e.g., geology, minerals, earthquakes, landslides, volcanoes, soils, gravity, permafrost, hydrogeology, erosion

health, 009

health, health services, human ecology, and safety

e.g., disease, illness, factors affecting health, hygiene, substance abuse

imageryBaseMapsEarthCover, 010

base maps

e.g., land cover, topographic maps, imagery, annotations

intelligenceMilitary, 011

military bases, structures, activities

e.g., military bases, structures, activities

inlandWaters, 012

inland water features, drainage systems and characteristics

e.g., rivers, glaciers, lakes, water use plans, dams, currents, floods, water quality, hydrographic charts

location, 013

positional information and services

e.g., addresses, geodetic networks, control points, postal zones, place names

oceans, 014

features and characteristics of salt water bodies

e.g., tides, tidal waves, coastal information, reefs

planningCadastre, 015

information used for appropriate actions for future use of the land

e.g., land use maps, zoning maps, cadastral surveys, land ownership

society, 016

characteristics of society and culture

e.g., anthropology, archaeology, religion, demographics, crime and justice

structure, 017

man-made construction

e.g., architecture, buildings, museums, churches, factories, housing, monuments, shops, towers

transportation, 018

means and aids for conveying persons and/or goods

e.g., roads, airports, airstrips, shipping routes, tunnels, nautical charts, vehicle and vessel locations, aeronautical charts, railways, trails

utilitiesCommunication, 019

energy, water and waste systems, and communications infrastructure

e.g., hydroelectricity, geothermal, solar, and nuclear sources of energy, water purification and distribution, sewage collection and disposal, electrical and gas distribution, data communication, telecommunication, radio, communication networks