

Metadata

GIS Applications

Spring 2009

What is Metadata?

Data about data

Any information that makes data useful for another user

Background information that describes source, content, quality, condition, availability, use conditions, and distribution methods for data.

Huh?

WQPW-ID	DIN	Pb
PB-31	.34	.012
HK-14	.12	.023
PB-12	.35	.034
WA-3	.28	.001
PB-4	.23	.022
PB-5	.21	.013

Why is it Important?

Supports data sharing

Helps users to find data

Helps users to judge quality/utility of the data

Extends useful life-span of data

Saves time and money

Why is it Important?

Helps maintain an organization's investment in spatial data.

Provides an inventory of data assets

Helps determine and maintain the value of data

Helps determine the reliability and currency of data

Supports decision making

Documents legal issues

Helps keep data accurate and helps verify accuracy

Helps determine budgets -when or if data needs to be updated or repurchased

Roles of Metadata

Supports access (search, browse, and retrieval)

Supports transfer

Supports evaluation of fitness for use

Supports use

Context of Use

Catalogues

Management records

Accompanying a dataset

Content Standard for Digital Geospatial Metadata

Federal Geographic Data Committee (FGDC) the adopted content standard for metadata in 1992 and revised it in 1998 Vers. 2 (FGDC-STD-001-1998).

Executive Order 12096 was signed by President Clinton on April 11, 1994

Required all US Federal agencies to use this standard in documenting newly created geospatial data as of January 1995.

The standard has been implemented beyond the federal level with State and local governments adopting the metadata standard as well.

Major Sections of the Content Standard

Identification Information

data set title, area covered, keywords, purpose, abstract, access and use restrictions

Data Quality Information

horizontal and vertical accuracy assessment, data set completeness and lineage

Spatial Data Organization Information

raster, vector, indirect geo-referencing

Spatial Reference Information

lat/long, coordinate system, map projection

Major Sections of the Content Standard

Entity and Attribute Information

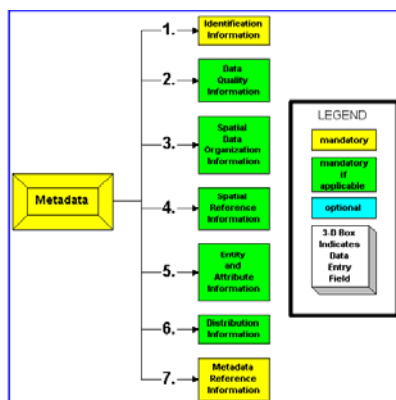
definitions of the attributes of the data set

Distribution Information

distributor, file format of the data, off line media type, on line link to data, fees

Metadata Reference Information

who created the metadata and when



Other Standards Efforts

US MARC – Machine readable cataloging

Dublin Core

GILS - Government Information Locator Service

<http://www.gpoaccess.gov/gils/index.html>

DIF - Directory Interchange Format (NASA)

ISO 19115 - International Standards Organization

Dublin Core

Subject	Identifier
Title	Relation
Author	Source
Publisher	Language
Other Agent	Coverage
Date	Minimum metadata elements for descriptions and discovery of web based documents
Object Type	
Form	

CSDGM Examples

Citation Information

Originator: Municipality of Murrysville community Development Office

Title: Murrysville Parcel Coverage

Abstract: This layer includes the parcels within the municipality along with condominiums and mobile homes. Also included are the county's tax data along with municipal data such as street address, lot number, and subdivision, build out status, land use zoning, septic or sewer and housing starts on two year intervals

Purpose: For community planning and demonstrations purposes

CSDGM Examples

Citation Information

Originator: PA Department of Environmental Protection

Title: Ambient and Fixed Station Network Groundwater Monitoring Point Data (1985-1997)

Abstract: This coverage represents the point locations and data for 1,089 groundwater quality monitoring points sampled under the PA DEP Fixed Station Network and Ambient Survey groundwater monitoring program. Sample data were collected from 1985-1997. Monitoring points were typically homeowner wells, springs, public water supplies, or industrial wells.

Examples - Keywords

Keywords:

Theme:

Theme Keyword: Wetlands

Theme Keyword: Arc/Info Coverage

Theme Keyword: Freshwater Wetlands

Theme Keyword: Regulatory

Place:

Place Keyword: Albany County

Place Keyword: New York

Examples - Distribution Information

[Online File Transfer \(ftp\)](#) (file size = 817819b) [Tips for downloading](#)

Distribution Liability: The USER shall indemnify, save harmless, and if requested defend those parties involved with the development and distribution of this data, their officers, agents, and employees from and against any suits, claims, or actions for injury, death, or property damage arising out of the use of or any defect in the FILES or accompanying documentation.

ISO 19115

ISO 19115 is the International Organization of Standards (ISO) approved international metadata standard.

As a member of ISO, the US is required to revise the CSDGM in accord with ISO 19115

ISO 19115 Core Metadata Elements

Mandatory Elements:

Dataset title
Dataset reference date
Dataset language
Dataset topic category
Abstract
Metadata point of contact
Metadata date stamp

Conditional Elements:

Dataset responsible party
Geographic location by coordinates
Dataset character set
Spatial resolution
Distribution format
Spatial representation type
Reference system
Lineage statement
On-line Resource
Metadata file identifier
Metadata standard name
Metadata standard version
Metadata language
Metadata character set

Objectives when creating Metadata

Create enough metadata so that any user within an organization can make effective use of a data set.

Create metadata compliant with the federal metadata standard so data can be shared outside the agency

Where Does One Begin?

What What data needs to be documented?

Who Who should do the documentation?

When When should it be compiled?

What data sets need to be documented?

Inventory data sets.

What data sets is the agency responsible for producing?

Prioritize data sets

Which are most important? Greatest current or anticipated future use. Basis for other data sets. Most expensive to collect.

Which are easiest to document?

Who should create the metadata?

Person most familiar with the data?

Person most familiar with metadata production?

How much can be generated automatically?

When should metadata be compiled?

Prior to data collection

Concurrently with data collection

After data collection

Need to keep metadata up-to-date

How to create metadata

Assume creation of CSDGM compliant metadata.

Understand the data and the standard

Review some examples

Use metadata tools

Validate the metadata

Categories of Metadata Tools

Intelligent - self extracting from a data set

Forms Based - guided process with pick lists

ASCII word processor templates - cut and paste

Utilities - finding, formatting, and validation tools

Example Metadata Tools

Freeware/Shareware tools:

- **Isme** :- Windows based tool developed and actively maintained by Peter Schweitzer of the USGS Geology Discipline. Supports the biological, shoreline, and remote sensing profiles/extensions.
- **xsmc** :- Unix based version of Isme.
- **MetaScript** :- Online utility for creating a custom metadata entry form (templates) for the collection of highly similar metadata from individual or multiple participants. Metadata conformance is checked upon submission using rpm. Developed and maintained by the NOAA Coastal Services Center.
- **MERMAID** :- Online utility for creating, editing, validating, storing and exporting metadata records via the NOAA Coastal Data Development Center. Provides rigorous validation, managed workflow with customized permissions, change tracking, and storage of files associated with the metadata. MERMAID supports the FGDC standard, the biological and shoreline profiles and the remote sensing extension, and provides the ability for metadata conversion among these standards. An export option includes the conversion of FGDC metadata to MARC-XML. MERMAID also supports the Ecological Metadata Language.
- **Metaview 2005** :- Stand-alone metadata creation and editing tool developed by the USDA Forest Service North Central Research Station. Metaview 2005 creates metadata compliant with the CSDGM (FGDC) 1998 metadata standard and the National Biological Information Infrastructure (NBII) 1999 Biological Data Profile for the FGDC standard. The software runs under the Microsoft Windows 2000 and XP operating systems, and requires the presence of Microsoft's .Net Framework version 1.1. The metadata are output in XML format.
- **EPA Metadata Editor** :- The EPA Metadata Editor (EME) is an extension to ESRI's ArcCatalog version 9.2. It allows users to efficiently create and edit geospatial metadata records that meet EPA's Geospatial Metadata Technical Specification Version 1.0 and Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) version 2.0 requirements.

Commercial tools:

- **ArcCatalog** :- ESRI's ArcGIS internal data management tool that includes metadata creation and editing functions, data search and discovery functions, auto-capture of data properties from ESRI data, auto-updates to metadata when data is edited (synchronization), bundling of metadata with data upon export, and the ability to create metadata in CSDGM and ISO (not US Profile) formats.

<http://www.fgdc.gov/metadata/geospatial-metadata-tools>

Metadata Formats

ASCII Text

HTML (Hypertext Markup Language)

XML (Extensible Markup Language)

Granularity of Metadata

Sets - Current focus

Supersets

Subsets

Feature level

Need an approach for handling inheritance in entry and updates

Metadata Summary

Metadata creation should be part of standard operating procedures.

Time and effort related to metadata should be entered into the budget or project plan of every GIS operation.

Metadata should be a fundamental component of every GIS implementation