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.

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

<table>
<thead>
<tr>
<th>WAQPW-ID</th>
<th>DIN</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB-31</td>
<td>34</td>
<td>012</td>
</tr>
<tr>
<td>HK-14</td>
<td>12</td>
<td>023</td>
</tr>
<tr>
<td>PB-12</td>
<td>35</td>
<td>034</td>
</tr>
<tr>
<td>WA-3</td>
<td>28</td>
<td>001</td>
</tr>
<tr>
<td>PB-4</td>
<td>23</td>
<td>022</td>
</tr>
<tr>
<td>PB-5</td>
<td>21</td>
<td>013</td>
</tr>
</tbody>
</table>
Data are the most expensive components of a GIS. Metadata is a means of preserving the value of data investments. This is of particular significance to local and regional governments experiencing rapid staff changes.

Metadata contain information that can be used to quickly locate and retrieve data resources by specific criteria including: keywords, time period, contacts, data type, entities and attributes, etc.

Metadata is increasingly used by software systems as a means of properly ingesting data and by analysts as a means of properly displaying data.

**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 data transfer
- Supports evaluation of fitness for use
- Supports use

**Metadata and Data Liability**
A well-written metadata record is an opportunity to state what the data are not.

Metadata can be useful in advising others in the appropriate and misappropriate application of the data.

An explicit purpose statement can clearly outline special project conditions and requirements that may affect the applicability of the data to other projects. Use constraint statements can be crafted to express scale, geographic, or temporal limitations to the data. Liability statements should be written by legal staff to ensure that the legal requirements for use of the data are fully outlined.
Metadata and Contract Deliverables

Metadata should be specified as a deliverable when contracting with others for the development of data.

The metadata specification should include clear language as to the metadata standard that should be used and provide some indication as to the quality of the metadata expected.

Context of Use

Catalogues

Management records

Accompanying a dataset

Content Standard for Digital Geospatial Metadata

The Federal Geographic Data Committee (FGDC) approved the first version of the Content Standard for Digital Geospatial Metadata (CSDGM) in June 1994.

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.

Content Standard for Digital Geospatial Metadata

In 1998, the CSDGM was revised and Version 2.0 was published.

In 1999 the International Standards Organization (ISO) Technical Committee (TC) 211 Geographic Information / Geomatics was tasked with harmonizing the CSDGM with other geospatial metadata standards and a range of de facto standards that had emerged to address new requirements for geospatial documentation.

The result was ISO 19115: Geographic Information – Metadata.
ISO 19115: Geographic Information – Metadata.


Since then, individual organizations and nations have developed implementation ‘profiles’ of the standard.

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

**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)

Dublin Core

<table>
<thead>
<tr>
<th>Subject</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Relation</td>
</tr>
<tr>
<td>Author</td>
<td>Source</td>
</tr>
<tr>
<td>Publisher</td>
<td>Language</td>
</tr>
<tr>
<td>Other Agent</td>
<td>Coverage</td>
</tr>
<tr>
<td>Date</td>
<td>Minimum metadata elements for descriptions and discovery of web based documents</td>
</tr>
<tr>
<td>Object Type</td>
<td></td>
</tr>
<tr>
<td>Form</td>
<td></td>
</tr>
</tbody>
</table>

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

ISO 19115


The effort to develop the NAP began through a Memorandum of Understanding (MoU) between INCITS, a standards development organization accredited by ANSI, and the Canadian General Standards Board – Council on Geomatics (CGSB-COG), a federal government organization and standards development organization accredited by the Standards Council of Canada (SCC).

In response to this MoU, U.S. and Canadian experts tailored ISO 19115:2003 to meet the requirements of both countries

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

Why Change Standards?

ISO 19115 development was initiated to:

Provide an internationally standardized means of documenting geospatial data resources

Incorporate international references including language and character sets

Address new geospatial data structures and models

Include geospatial data applications and services.
As a member of ISO, the US is required to revise the CSDGM in accord with ISO 19115

ISO 19115 Core Metadata Elements

<table>
<thead>
<tr>
<th>Mandatory Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset title</td>
</tr>
<tr>
<td>Dataset reference date</td>
</tr>
<tr>
<td>Dataset language</td>
</tr>
<tr>
<td>Dataset topic category</td>
</tr>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Metadata point of contact</td>
</tr>
<tr>
<td>Metadata date stamp</td>
</tr>
</tbody>
</table>

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 metadata standards so data can be shared outside the agency or organization.

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?
Prioritize data sets

Prioritize the order in which data sets are documented based upon the following:

- value of data set as a core, or framework, data product of the organization.
- utility of the data set within the organization
- number of external requests for the data set
- historical significance of the data set to the organization

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 standards 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

Example 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

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