Introduction to Geospatial Metadata

D G Rossiter
Cornell University, Section of Soil & Crop Sciences
南京师范大学地理学学院

January 9, 2018

What

Metadata: structured information that describes an information resource: "data about data"

- **structure**: how it is organized, how to access
- **content**: what it represents
- **lineage**: how it was created

Digital metadata: machine-readable (allows **data discovery**)

Geospatial metadata

Metadata about objects with a geographic **location** and **extent**

i.e., somewhere “on” the Earth’s suface (or referenced to it).

Why

- locate (data discovery)
- describe
- use (structure, meaning)
- administer (access rights etc.)
- preserve (archive) for eventual re-use

Standards

ISO 19915: Geographic Information – Metadata

http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html

Content standard for digital geospatial metadata (FGDC-STD-001-1998)

http://www.fgdc.gov/metadata/us-national-profile-iso19115/index_html

Explains how FGDC will adjust this into The North American Profile (NAP) [USA and Canada] of ISO 19115

http://geology.usgs.gov/tools/metadata/tools/doc/plain.faq.html

In plain language

What does this data set describe?
- How should this data set be cited?
- What geographic area does the data set cover?
- Does the data set describe conditions during a particular time period?
- What is the general form of this data set?
- How does the data set represent geographic features?
- How does the data set describe geographic features?

Who produced the data set?
- Who are the originators of/contributors to the data set?
- To whom should users address questions about the data?

Why was the data set created?

How was the data set created?
- Where did the data come from?
- What changes have been made?

How reliable are the data; what problems remain in the data set?
- How well have the observations been checked?
- How accurate are the geographic locations, and heights or depths (if applicable)?
- Where are the gaps in the data? What is missing?
- How consistent are the relationships among the data, including topology?

How can someone get a copy of the data set?
- Are there legal restrictions on access or use of the data?
- Who distributes the data?
- What's the catalog number I need to order this data set?
- What legal disclaimers am I supposed to read?
- How can I download or order the data?

Who wrote the metadata?

More formally ...

Identification Information Basic information about the data set.

Examples: title, geographic area covered, currentness, rules for acquiring or using the data.

Data Quality Information An assessment of the quality of the data set.

Examples: positional and attribute accuracy, completeness, consistency, sources of information, and methods used to produce the data.

Spatial Data Organization Information The mechanism used to represent spatial information

Examples: direct (raster or vector), indirect (street addresses, county codes); the number of spatial objects in the data set.

Spatial Reference Information The reference frame for, and means of encoding, coordinates

Examples: name and parameters for map projections or grid coordinate systems, horizontal and vertical datum, coordinate system resolution.

Entity and Attribute Information The entity types and their attributes and the domains from which attribute values may be assigned.

Examples: names and definitions of features, attributes, and attribute values; reference to data dictionary

Distribution Information How to obtain the data set

Examples: contact for the distributor, available formats, online or physical media access, and fees for the data.

Metadata Reference Information Which metadata standard, who wrote this metadata?

HTML metadata

Tompkins County Agricultural Districts

Identification_Information:
Citation:
Citation_Information:
Originator: Cornell Institute for Resource Information Sciences (Cornell IRIS)
Publication_Date: 20140228
Title: Tompkins County Agricultural Districts
Geospatial_Data_Presentation_Form: vector digital data
Online_Linkage: http://cugir.mannlib.cornell.edu/bucketinfo.jsp?id=7994
Description:
Abstract: These GIS files represent geographic boundaries for lands that are under the protection of NYS Agricultural District Law, administered by the New York State Department of Agriculture and Markets. The boundaries are derived from New York State Agricultural District, 1:24,000-scale, maps produced at county agencies. The district boundaries correspond to tax parcel data. District boundaries are joined into a file representing all of the Agricultural Districts within an entire county. Note that 2003 legislation allows lands to be added to districts on an annual basis. Electronic data provided here may predate those additions. Tax parcel detail and secondary rights-of-way are not included in this dataset. Rights-of-way for state and federal highways, railroads and utilities are only included when they are delineated on the original 1:24,000 scale maps. The data files are in ArcGIS shapefile format.

HTML metadata - spatial reference information

Spatial_Reference_Information:
Horizontal_Coordinate_System_Definition:
Planar:
Grid_Coordinate_System:
Grid_Coordinate_System_Name: Universal Transverse Mercator
Universal_Transverse_Mercator:
UTM_Zone_Number: 18
Transverse_Mercator:
Scale_Factor_at_Central_Meridian: 0.9996
Longitude_of_Central_Meridian: -75
Latitude_of_Projection_Origin: 0
False_Easting: 500000
False_Northing: 0
Planar_Coordinate_Information:
Planar_Coordinate_Encoding_Method: Coordinate pair
Coordinate_Representation:
Abscissa_Resolution: 1
Ordinate_Resolution: 1
Planar_Distance_Units: Meters
Geodetic_Model:
Horizontal_Datum_Name: North American Datum of 1983
Ellipsoid_Name: GRS-80
Semi-major_Axis: 6378137
Denominator_of_Flattening_Ratio: 298.257222

HTML metadata - attribute information

Entity_and_Attribute_Information:
Detailed_Description:
Entity_Type:
Entity_Type_Label: tomp_lulc_2012
Entity_Type_Definition: Land use or land cover class, 2012
Entity_Type_Definition_Source: Classification system is based on LUNR, but has been adapted to meet the needs of the Tompkins County Planning Department
Attribute:
Attribute_Label: OBJECTID
Attribute_Definition: Internal ID
Attribute_Definition_Source: Tompkins County Planning Department and Tompkins County ITS GIS Division
Attribute_Domain_Values:
Unrepresentable_Domain: Positive integers that are automatically generated.
Attribute:
Attribute_Label: LU
Attribute_Definition: Land Use and Land Cover classes devised by the staff in the Tompkins County Planning Department and the Tompkins County ITS GIS Division, based in part on the LUNR classification system from 1968.
Attribute_Definition_Source: Tompkins County Planning Department and Tompkins County ITS GIS Division, 1999
Attribute_Domain_Values:
Enumerated_Domain:
Enumerated_Domain_Value: Ac
Enumerated_Domain_Value_Definition: Cropland: Tillable land used for growing cultivated field crops, forage crops, grain, beans, etc. Hedgerows separating defined Ac areas were delineated as separate classes (typically Fd, Fm, Fb or Fc) if they are greater than 20 meters wide
Enumerated_Domain_Value_Definition_Source: Tompkins County Planning Department and Tompkins County ITS GIS Division, 1999.

How

1. Describe in plain language – but how do you know you've covered everything?

2. Use **formal metadata tools**

    - During data design
    - During data acquisition
    - At the end of the project

Metadata tools

Lists

- http://www.fgdc.gov/metadata/us-national-profile-iso19115/geospatial-metadata-tools

Useful tools

1. Creation

   - ArcCatalog (ESRI); integrated with ArcGIS.
   - https://edg.epa.gov/EME/ EPA metadata editor (requires Microsoft .net)
   - http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html
     Program: tkme

2. Checking, export

   - http://geology.usgs.gov/tools/metadata/tools/doc/mp.html
     Program: mp (Metadata Parser)
   - on-line validation (http://geo-nsdi.er.usgs.gov/validation/