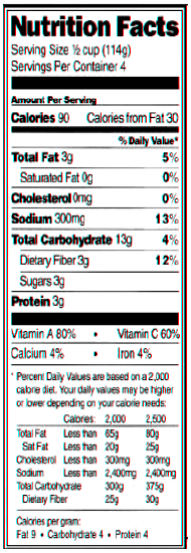
Describing datasets (2)

IST400/600

Jian Qin

An analogy



Nutrition Facts

Serving Size ½ cup (114g)
Servings Per Container 4

Amount Per Serving

Calories 90 Calories from Fat 30

	% Daily Value*
Total Fat 3g	5%
Saturated Fat 0g	0%
Cholesterol 0mg	0%
Sodium 300mg	13%
Total Carbohydrate 13g	4%
Dietary Fiber 3g	12%
Sugars 3g	
Protein 3g	

Vitamin A 80% • Vitamin C 60%
Calcium 4% • Iron 4%

* Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:

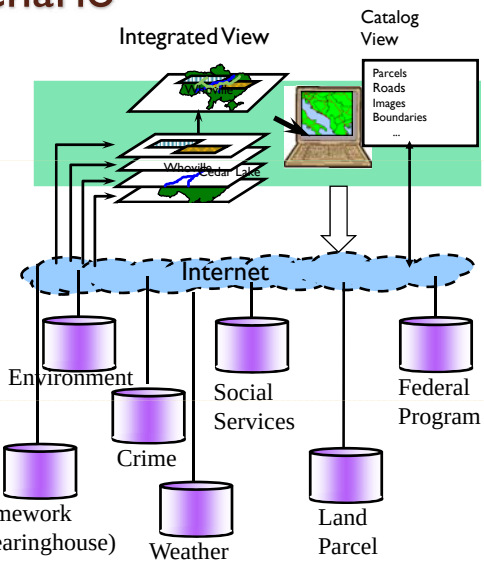
	Calories	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300mg	300mg
Sodium	Less than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

Calories per gram:
Fat 9 • Carbohydrate 4 • Protein 4

- What a nutritional label tells us about the food?
- Mandatory and voluntary components in the label

A use case scenario

- Discover, access, and retrieve data and views from multiple servers
- Query and update
- Provide interface for visualization, decision support, etc.
- Support geocoded raster, vector, matrix, and other data



Source: http://geongrid.org/CSIG05/presentations/IlyaZaslavsky/4_webGIS_sources.ppt

Metadata for users

- Support various search options (similar to using a library catalog)
- Help users decide the fitness of a dataset for use
- Provide information for handle data effectively, e.g. format
- Assist in identifying notable data contents, e.g. find images of an interesting hurricane

Metadata is expensive to generate, but the benefits are substantial

Metadata for data management

- Preserve the data history so that it can be re-used or adapted,
- Assess the age and character of data holdings to determine which data should be maintained, updated, or deleted,
- Instill data accountability by requiring you to state what you know about the data and realizing what you don't, but should, know about your data
- Limit data liability by explicitly designating the effective and administrative limits of use of the data. View an example

Metadata for project management

- Plan and document the data types and content needed to support the project
- Monitor data development by regular review of the process steps completed and recorded within the metadata
- Provide all project participants a common language of attributes and process methods and a place to record and share their progress
- Access the lineage and content of outsourced data production by requiring robust metadata as a contract deliverable.

Types of data standards

- Data Classification
  - e.g., Vegetation Classification
- Data Content
  - e.g., Digital Geospatial Metadata, Spatial Schema
- Data Symbology or Presentation
  - e.g., Digital Geologic Map Symbolization
- Data Transfer
- Data Services (Web Mapping, Feature)
- Data Usability
  - e.g., Geospatial Positioning Accuracy

The U.S. FGDC metadata standard

Content Standard for Digital Geospatial Metadata (CSDGM)

Extensions

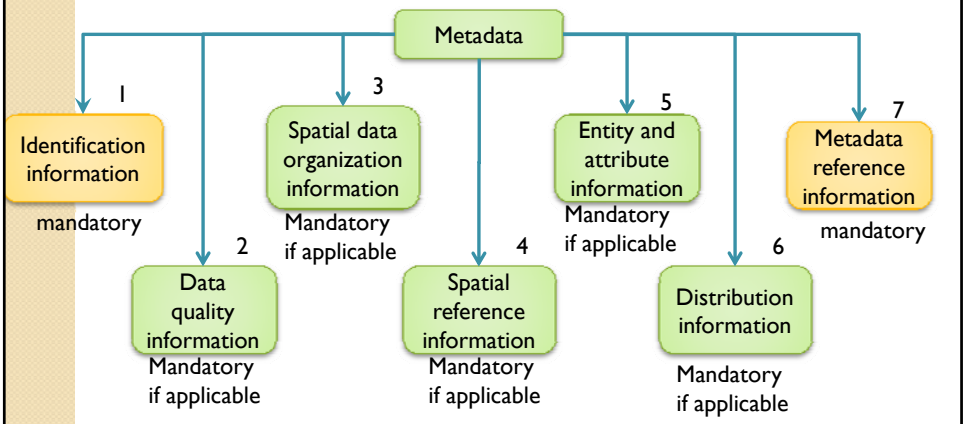
Extensions for Remote Sensing Metadata

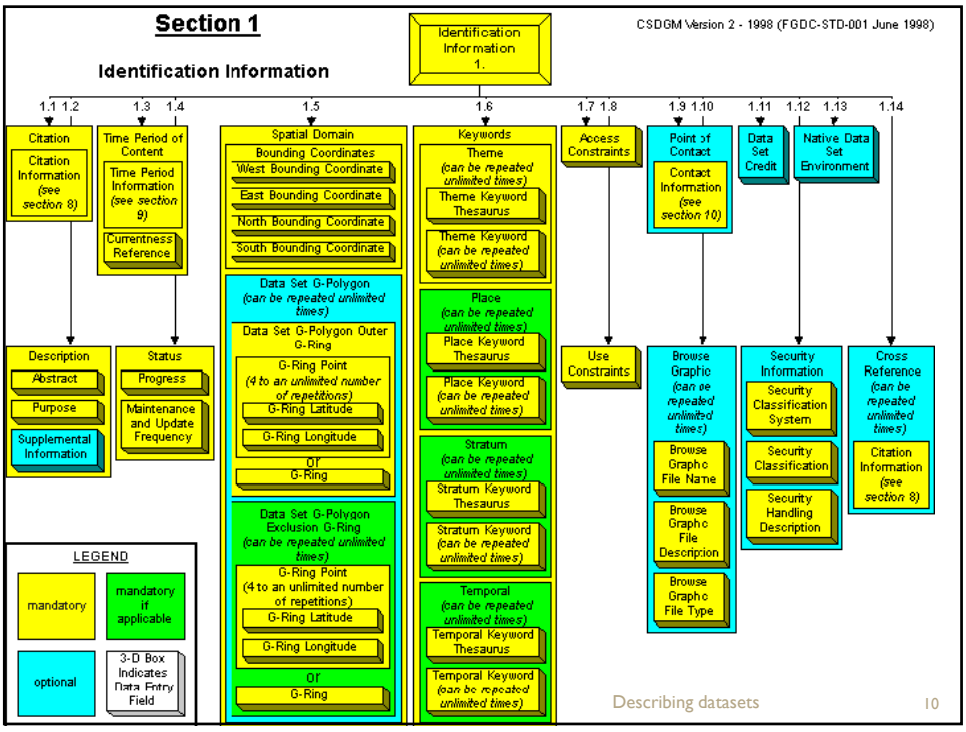
Biological Data Profile

Metadata Profile for Shoreline Data

Content Standard for Digital Geospatial Metadata (CSDGM)

- Sections
- Compound elements
- Data elements





Data element

- Form:

Data element name -- definition.

Type: (choice of "integer", "real", "text", "date", or "time")

Domain: (describes valid values that can be assigned)

- Example:

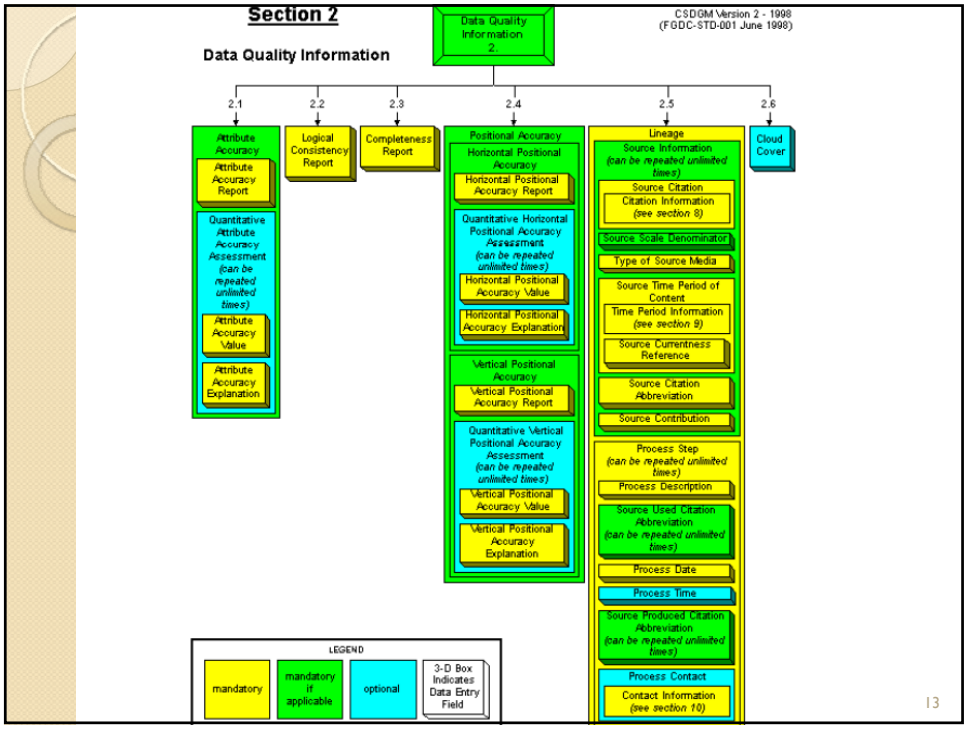
Abstract -- a brief narrative summary of the data set.

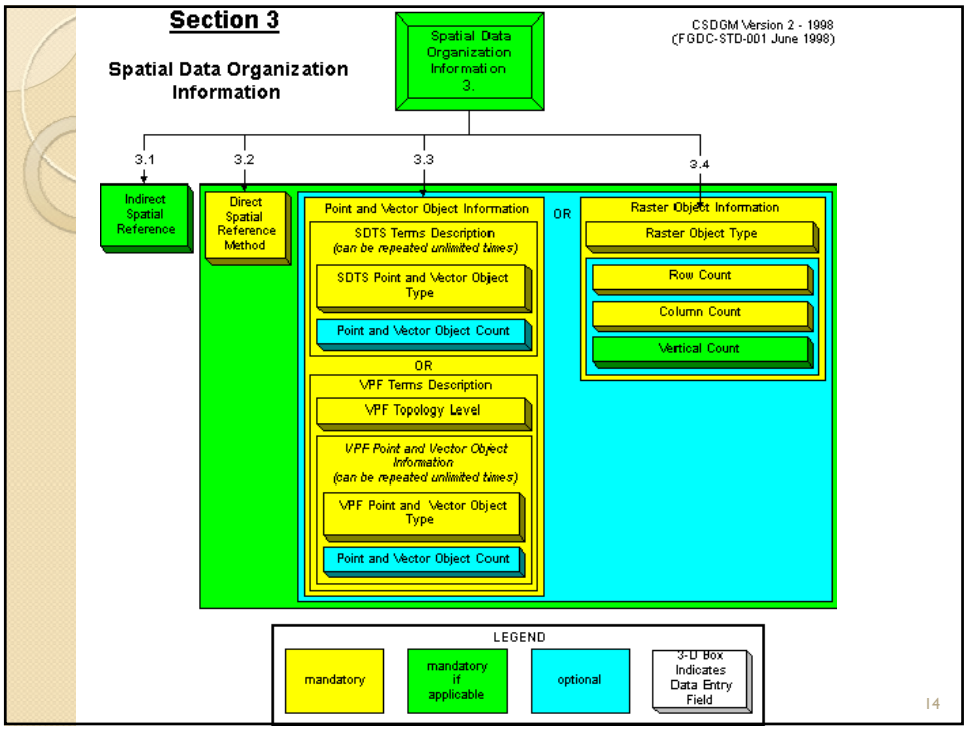
Type: text

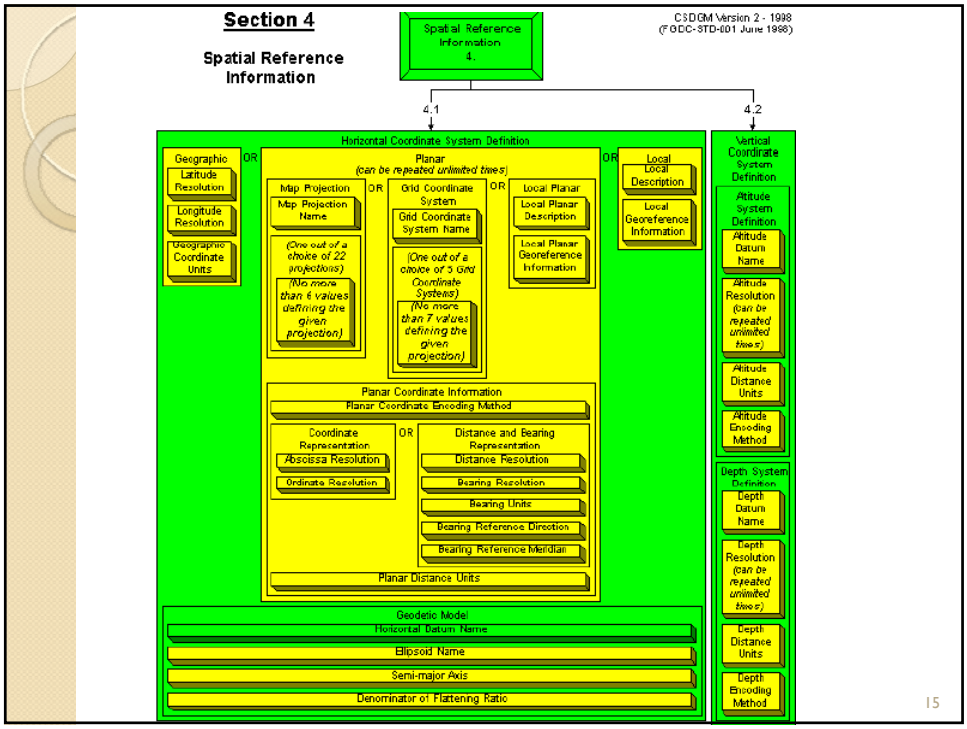
Domain: free text

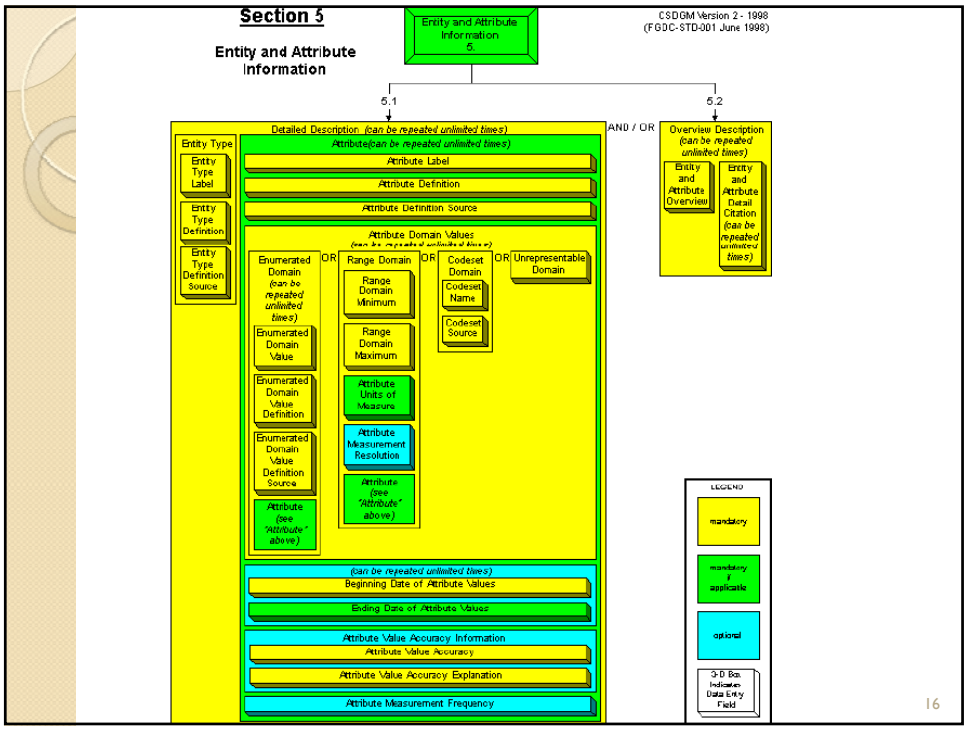
Data element: values

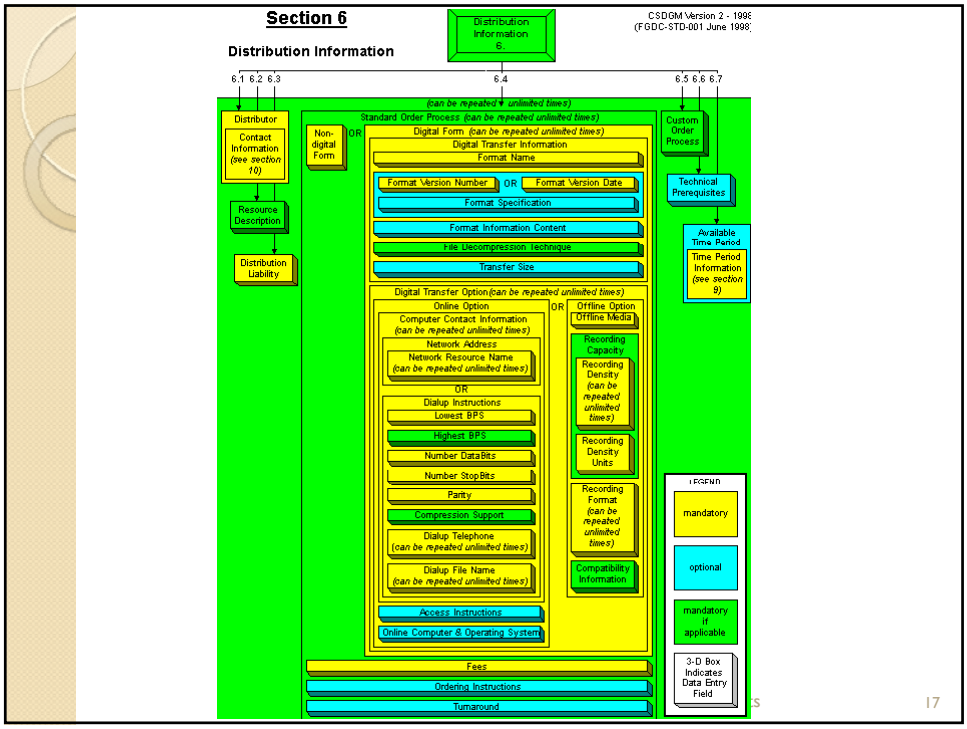
- The **domain** of values for a data element:
  - may be specified only by type. In these cases, the domain is described by the word "free" followed by the type (for example, free integer, free text).
  - may be specified by a list of values, references to a list of values, or a range of values.
  - may be partly specified by a set of values, followed by the "free" convention. In this case, values should be selected from the provided set; if no value is adequate, the user may provide one.

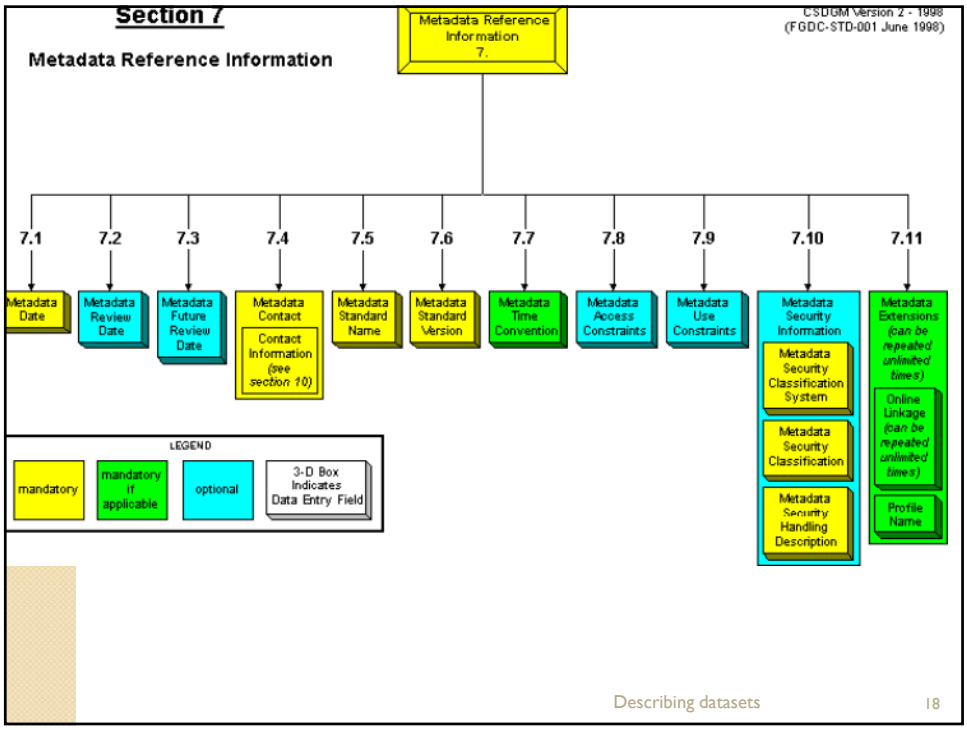

Section 2
Data Quality Information
CSDGM Version 2 - 1998
(FGDC-STD-001 June 1998)
Data Quality Information
2.
2.1
2.2
2.3
2.4
2.5
2.6
Attribute Accuracy
Attribute Accuracy Report
Quantitative Attribute Accuracy Assessment (can be repeated unlimited times)
Attribute Accuracy Value
Attribute Accuracy Explanation
Logical Consistency Report
Completeness Report
Positional Accuracy
Horizontal Positional Accuracy
Horizontal Positional Accuracy Report
Quantitative Horizontal Positional Accuracy Assessment (can be repeated unlimited times)
Horizontal Positional Accuracy Value
Horizontal Positional Accuracy Explanation
Vertical Positional Accuracy
Vertical Positional Accuracy Report
Quantitative Vertical Positional Accuracy Assessment (can be repeated unlimited times)
Vertical Positional Accuracy Value
Vertical Positional Accuracy Explanation
Lineage
Source Information (can be repeated unlimited times)
Source Citation
Citation Information (see section 8)
Source Scale Denominator
Type of Source Media
Source Time Period of Content
Time Period Information (see section 9)
Source Currentness Reference
Source Citation Abbreviation
Source Contribution
Process Step (can be repeated unlimited times)
Process Description
Source Used Citation Abbreviation (can be repeated unlimited times)
Process Date
Process Time
Source Produced Citation Abbreviation (can be repeated unlimited times)
Process Contact
Contact Information (see section 10)
Cloud Cover
LEGEND
mandatory
mandatory if applicable
optional
3-D Box Indicates Data Entry Field
13


Section 3
Spatial Data Organization Information
CSDGM Version 2 - 1998
(FGDC-STD-001 June 1998)
Spatial Data Organization Information
3.
3.1
3.2
3.3
3.4
Indirect Spatial Reference
Direct Spatial Reference Method
Point and Vector Object Information
SDTS Terms Description (can be repeated unlimited times)
SDTS Point and Vector Object Type
Point and Vector Object Count
OR
VPF Terms Description
VPF Topology Level
VPF Point and Vector Object Information (can be repeated unlimited times)
VPF Point and Vector Object Type
Point and Vector Object Count
OR
Raster Object Information
Raster Object Type
Row Count
Column Count
Vertical Count
LEGEND
mandatory
mandatory if applicable
optional
3-D Box Indicates Data Entry Field
14


Section 4
Spatial Reference Information
Spatial Reference Information 4.
CSDGM Version 2 - 1998 (FGDC-STD-001 June 1998)
4.1
4.2
Horizontal Coordinate System Definition
Geographic
Latitude Resolution
Longitude Resolution
Geographic Coordinate Units
OR
Planar (can be repeated unlimited times)
Map Projection
Map Projection Name
(One out of a choice of 22 projections)
(No more than 6 values defining the given projection)
Grid Coordinate System
Grid Coordinate System Name
(One out of a choice of 5 Grid Coordinate Systems)
(No more than 7 values defining the given projection)
Local Planar
Local Planar Description
Local Planar Georeference Information
Planar Coordinate Information
Planar Coordinate Encoding Method
Coordinate Representation
Abscissa Resolution
Ordinate Resolution
Distance and Bearing Representation
Distance Resolution
Bearing Resolution
Bearing Units
Bearing Reference Direction
Bearing Reference Meridian
Planar Distance Units
Local
Local Description
Local Georeference Information
Geodetic Model
Horizontal Datum Name
Ellipsoid Name
Semi-major Axis
Denominator of Flattening Ratio
Vertical Coordinate System Definition
Altitude System Definition
Altitude Datum Name
Altitude Resolution (can be repeated unlimited times)
Altitude Distance Units
Altitude Encoding Method
Depth System Definition
Depth Datum Name
Depth Resolution (can be repeated unlimited times)
Depth Distance Units
Depth Encoding Method
15


Section 5
Entity and Attribute Information
Entity and Attribute Information 5.
CSDGM Version 2 - 1998 (FGDC-STD-001 June 1998)
5.1
5.2
AND / OR
Detailed Description (can be repeated unlimited times)
Entity Type
Entity Type Label
Entity Type Definition
Entity Type Definition Source
Attribute (can be repeated unlimited times)
Attribute Label
Attribute Definition
Attribute Definition Source
Attribute Domain Values (can be repeated unlimited times)
Enumerated Domain (can be repeated unlimited times)
Enumerated Domain Value
Enumerated Domain Value Definition
Enumerated Domain Value Definition Source
Attribute (see "Attribute" above)
Range Domain
Range Domain Minimum
Range Domain Maximum
Attribute Units of Measure
Attribute Measurement Resolution
Attribute (see "Attribute" above)
Codeset Domain
Codeset Name
Codeset Source
Unrepresentable Domain
(can be repeated unlimited times)
Beginning Date of Attribute Values
Ending Date of Attribute Values
Attribute Value Accuracy Information
Attribute Value Accuracy
Attribute Value Accuracy Explanation
Attribute Measurement Frequency
Overview Description (can be repeated unlimited times)
Entity and Attribute Overview
Entity and Attribute Detail Citation (can be repeated unlimited times)
LEGEND
mandatory
mandatory if applicable
optional
3-D Box Indicates Data Entry Field
16


Section 6
Distribution Information
Distribution Information 6.
CSDGM Version 2 - 1998 (FGDC-STD-001 June 1998)
6.1 6.2 6.3
6.4
6.5 6.6 6.7
(can be repeated unlimited times)
Distributor
Contact Information (see section 10)
Resource Description
Distribution Liability
Non-digital Form
OR
Standard Order Process (can be repeated unlimited times)
Digital Form (can be repeated unlimited times)
Digital Transfer Information
Format Name
Format Version Number
OR
Format Version Date
Format Specification
Format Information Content
File Decompression Technique
Transfer Size
Digital Transfer Option (can be repeated unlimited times)
Online Option
OR
Offline Option
Offline Media
Computer Contact Information (can be repeated unlimited times)
Network Address
Network Resource Name (can be repeated unlimited times)
OR
Dialup Instructions
Lowest BPS
Highest BPS
Number DataBits
Number StopBits
Parity
Compression Support
Dialup Telephone (can be repeated unlimited times)
Dialup File Name (can be repeated unlimited times)
Access Instructions
Online Computer & Operating System
Recording Capacity
Recording Density (can be repeated unlimited times)
Recording Density Units
Recording Format (can be repeated unlimited times)
Compatibility Information
Fees
Ordering Instructions
Turnaround
Custom Order Process
Technical Prerequisites
Available Time Period
Time Period Information (see section 9)
LEGEND
mandatory
optional
mandatory if applicable
3-D Box Indicates Data Entry Field
17


Section 7
Metadata Reference Information
Metadata Reference Information 7.
CSDGM Version 2 - 1998 (FGDC-STD-001 June 1998)
7.1
7.2
7.3
7.4
7.5
7.6
7.7
7.8
7.9
7.10
7.11
Metadata Date
Metadata Review Date
Metadata Future Review Date
Metadata Contact
Contact Information (see section 10)
Metadata Standard Name
Metadata Standard Version
Metadata Time Convention
Metadata Access Constraints
Metadata Use Constraints
Metadata Security Information
Metadata Security Classification System
Metadata Security Classification
Metadata Security Handling Description
Metadata Extensions (can be repeated unlimited times)
Online Linkage (can be repeated unlimited times)
Profile Name
LEGEND
mandatory
mandatory if applicable
optional
3-D Box Indicates Data Entry Field
Describing datasets
18

Implementation decisions

- What is the best time to collect metadata?
- Who is this for, anyway?
  - Management – catalog – transfer
  - Details, details
- Legacy data and the future

Using tools to create metadata records

- Online utility
  - MetaScribe

    http://www.csc.noaa.gov/cgi-bin/id/metadata/metascribe/metascribe.cgi

    Demo: example template file
- Stand alone
  - Metavist demo