Geospatial Metadata Development and Use

Lynda Wayne
FGDC Metadata Education Coordinator / GeoMaxim

Objectives

After the workshop, students can:

- effectively comprehend metadata
- make the business case for metadata
- locate and access available metadata resources
- implement metadata creation within their own organizations

What IS Metadata?

Data 'reporting'

- **WHO** created the data?
- **WHAT** is the content of the data?
- **WHEN** was it created?
- **WHERE** is it geographically?
- **HOW** was the data developed?
- **WHY** was the data developed?

What IS Metadata?

What IS Metadata?

entity

attributes

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

view actual metadata record

Let's Make Metadata

Turn to your neighbor and document the following:

Title (name)

Abstract (general description)

Supplemental Information (uniqueness)

Who Creates Metadata?

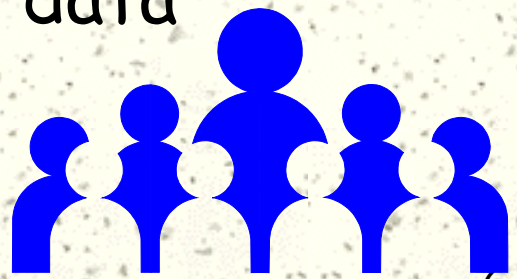
Officially

- Federal organizations producing digital geospatial data
- Federally funded projects that produce digital geospatial data

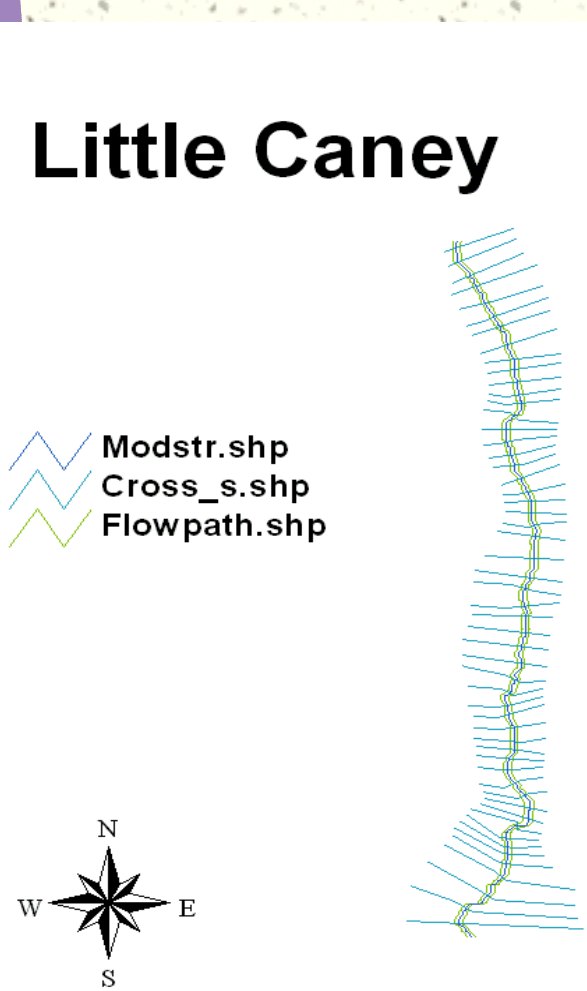
Morally and practically

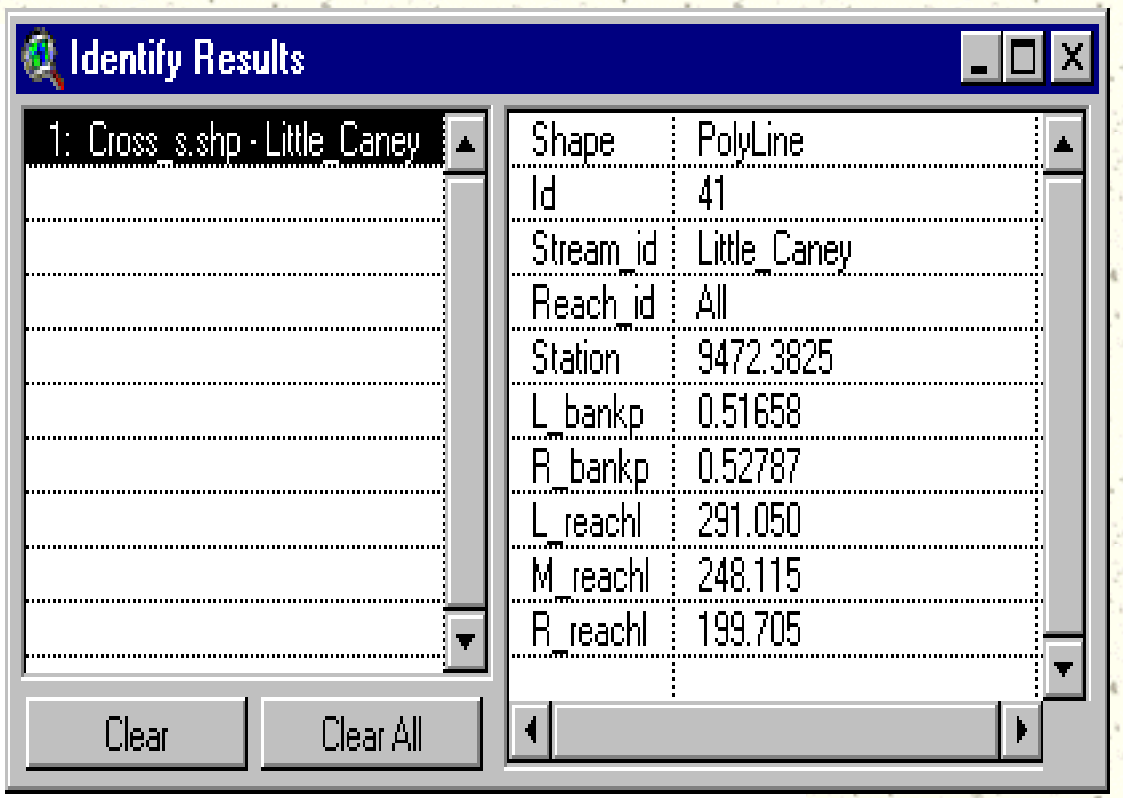
- Anybody that creates digital data

It's the right thing to do!



Why Create Metadata?



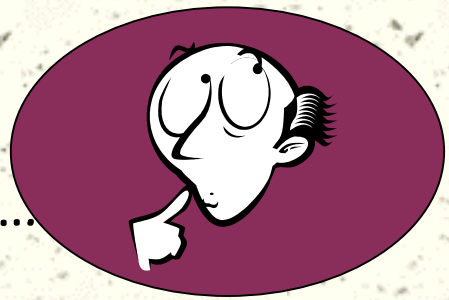


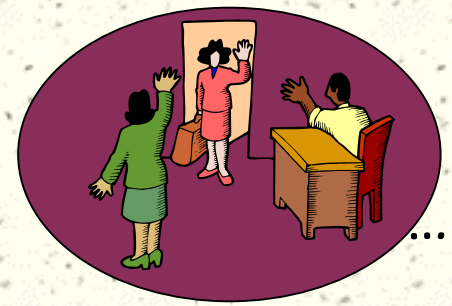
Why Create Metadata?

GIS Best Practice

- instill data accountability
- limit data liability
- preserve investments

people forget……





……leave

Why Create Metadata?

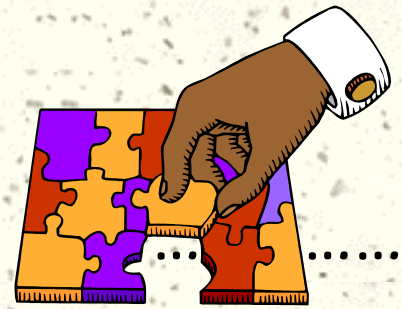
Responsible Community Member

- disclosure to citizens / clients
- information exchange with associates
- participation within the NSDI

Why Create Metadata?

National Spatial Data Infrastructure

Old perspective……



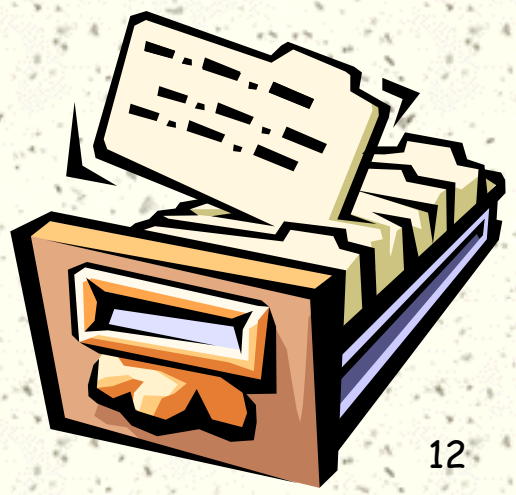
……New perspective

Why Create Metadata?

Effective Data Management

- locate your own data resources
- data maintenance and update
- locate external data resources

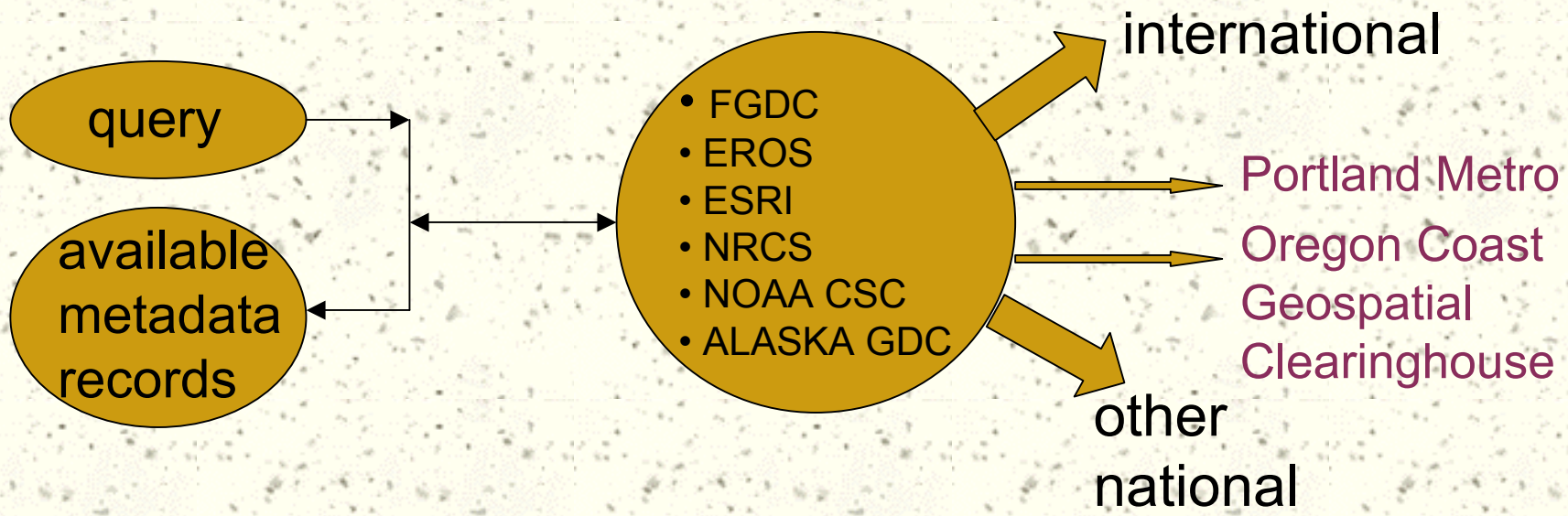
National Spatial Data Clearinghouse…



Why Create Metadata?

NSDI Geospatial Data Clearinghouse

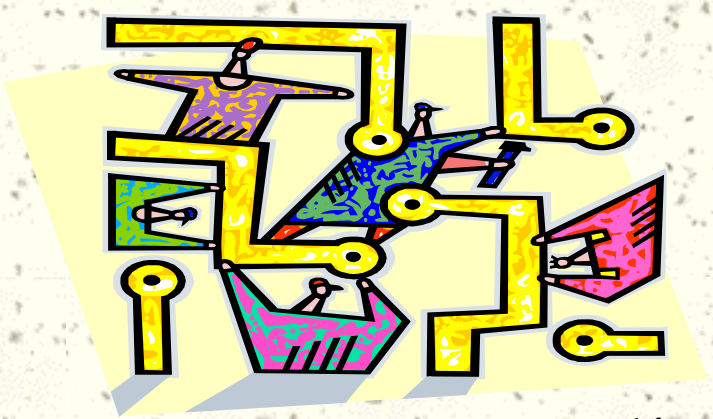
- international network of metadata distribution 'nodes'



Why Create Metadata?

Efficient Project Management

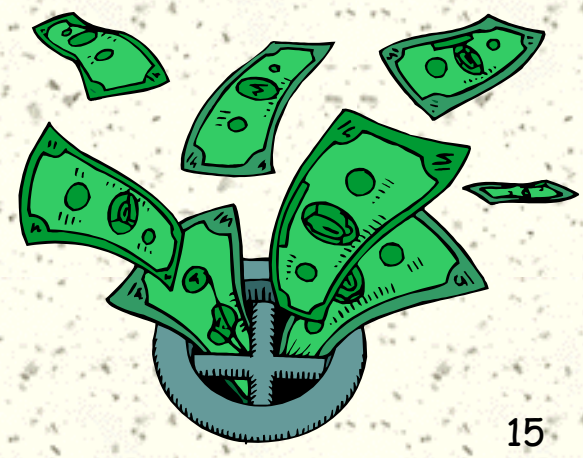
- monitor project development
- repeatable (defensible) process
- more accurate project estimates



Why Create Metadata?

Reduce Business Costs

- reduce data redundancy
- improve data utility
- assess data utility (CBA)
- fewer data inquiries



Why Create Metadata?

Enhance Security and Response

- faster access to data resource
- response characterization
- fitness of use assessment

…..the 9/11 experience

How Do We Get Started?

1. Select an appropriate metadata standard

Select a Metadata Standard

Dublin Core

- focus on core, minimal, elements
- bias toward online documents

FGDC CSDGM

- focus on robust data profiling
- bias toward digital geospatial data

CSDGM Profiles

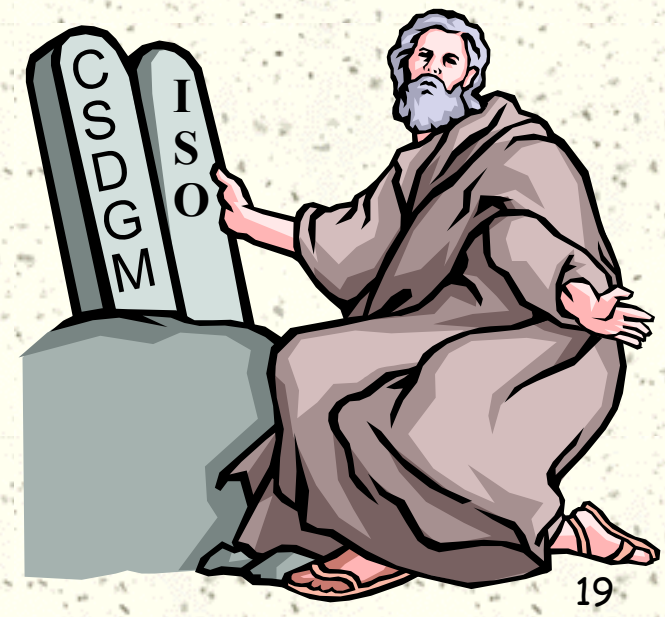
- bias toward disciplinary data

CSDGM

the metadata gospel…

Content Standard for Digital Geospatial Metadata

- It's not pretty
- It's not easy
- But sure is thorough…..

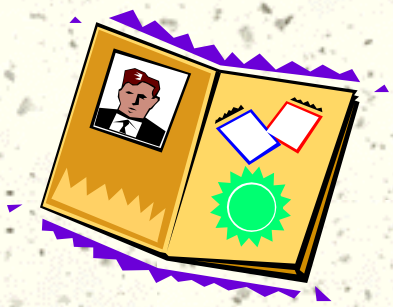


CSDGM

Seven Primary Sections

1. Identification Information
2. Data Quality Information
3. Spatial Data Organization Information
4. Spatial Reference Information
5. Entity and Attribute Information
6. Distribution Information
7. Metadata Reference Information

CSDGM



1. Identification Information

Information needed for data 'discovery'

General bibliographic information

Elements include:

- title, originator, data contact, status, date, time period of content, abstract, purpose, keywords, geographic location

CSDGM

2. Data Quality Information

How good is the data?

What is its lineage?

Did you 'check' the data?

- sources used, data processing methods and contacts, integrity checks, accuracy assessments

CSDGM

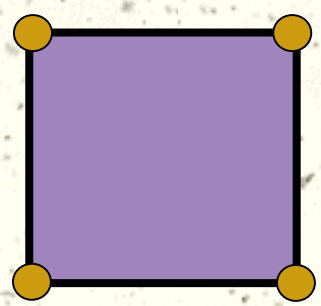
3. Spatial Data Organization

Data format:

- vector, point, raster

Indirect Spatial Reference:

- FIPS, PIN, PLSS



CSDGM

4. Spatial Reference Information

Coordinate system

- horizontal / vertical coordinate system, projection, datum

CSDGM

5. Entity and Attribute Information

Database design

entity

attributes

definitions

D = Divided

U = Undivided

PM = Planted Median

State Highways

year built	type	# lanes
1947	D	4
1973	PM	6
1975	PM	4
1980	U	2

domains

CSDGM

6. Distribution Information

How to acquire the data

- distribution contact, available formats, online distribution website, costs

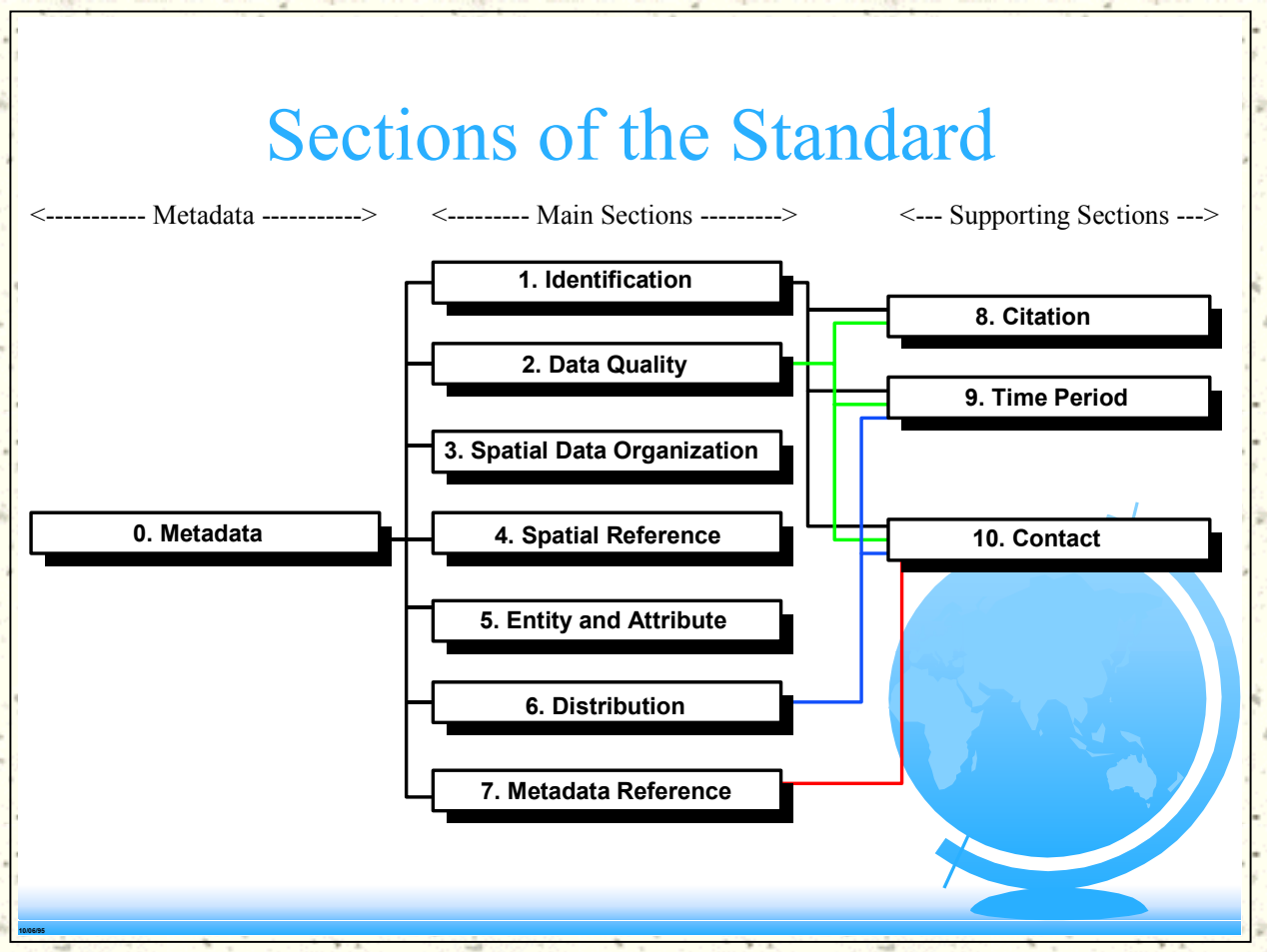
CSDGM

7. Metadata Reference Information

General information about the *metadata* record itself

- metadata contact, metadata standard used, metadata creation date, metadata review date

CSDGM


Sections of the Standard
<---------- Metadata ----------> <--------- Main Sections ---------> <--- Supporting Sections --->
0. Metadata
1. Identification
2. Data Quality
3. Spatial Data Organization
4. Spatial Reference
5. Entity and Attribute
6. Distribution
7. Metadata Reference
8. Citation
9. Time Period
10. Contact

CSDGM

Compound Element

A group of related data elements or other compound elements.

All compound elements ultimately are comprised of data elements.

Form:

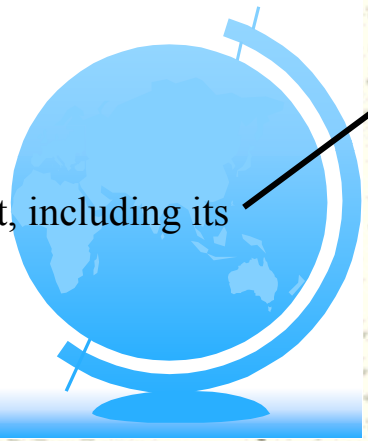
Compound element name -- definition.

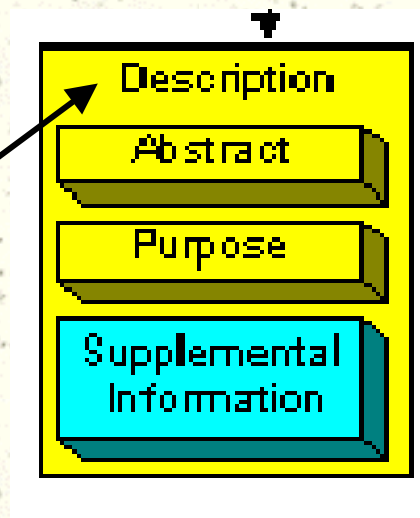
Type: compound

Example:

Description -- a characterization of the data set, including its intended use and limitations.

Type: compound





CSDGM

Data Element

A logically primitive item of data.

Data elements are the things that you "fill in."

Form:

Data element name -- definition.

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

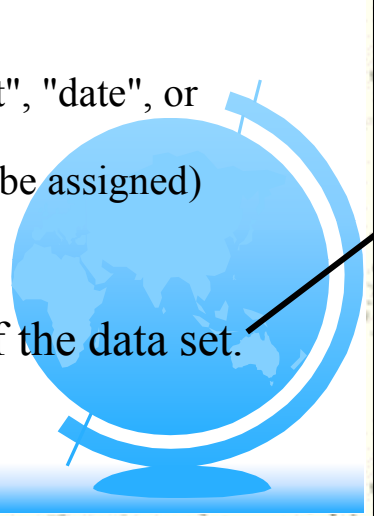
Domain: (describes valid values that can be assigned)

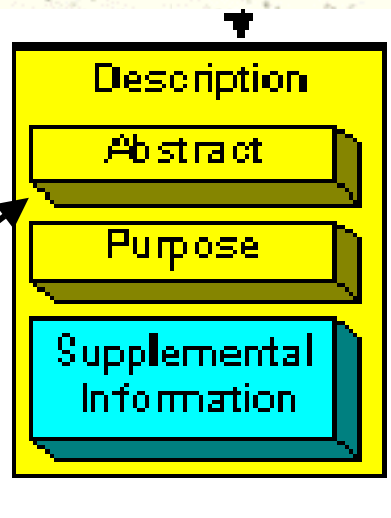
Example:

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

Type: text

Domain: free text





CSDGM

Data Element Forms for Special Values

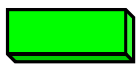
The standards specify the forms of four types of values:

- Calendar Dates
  - YYYYMMDD
- Time of Day
  - HHMMSSSS
- Coordinates
  - Lat/Lon Decimal Degrees
- Network addresses & file names
  - Service://hostname:port/path/filename

CSDGM

Production Rules

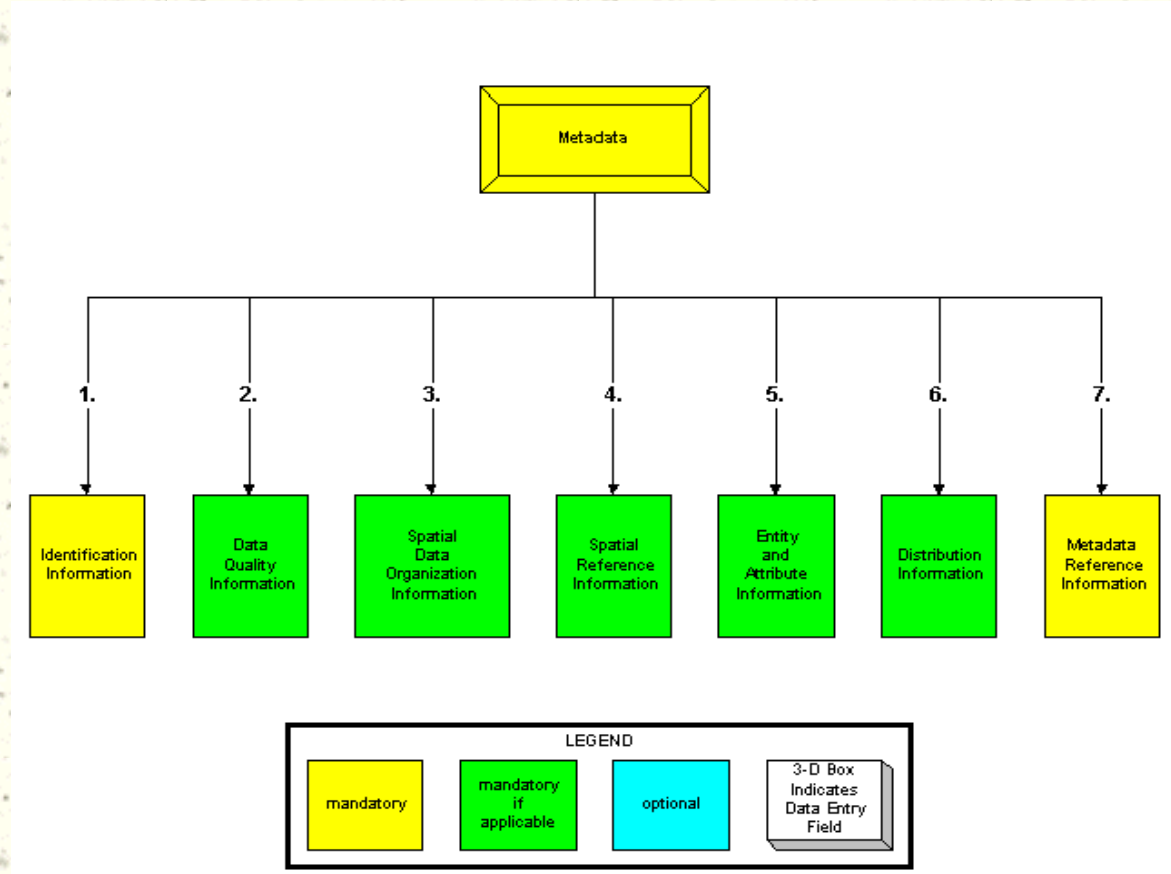
What's Mandatory? What's Not?

Compound Element	Data Element	Meaning
		**Mandatory - must be provided.**
		***Mandatory if Applicable - must be provided if the data set exhibits the defined characteristic.***
		**Optional - provided at the discretion of the data set producer.**

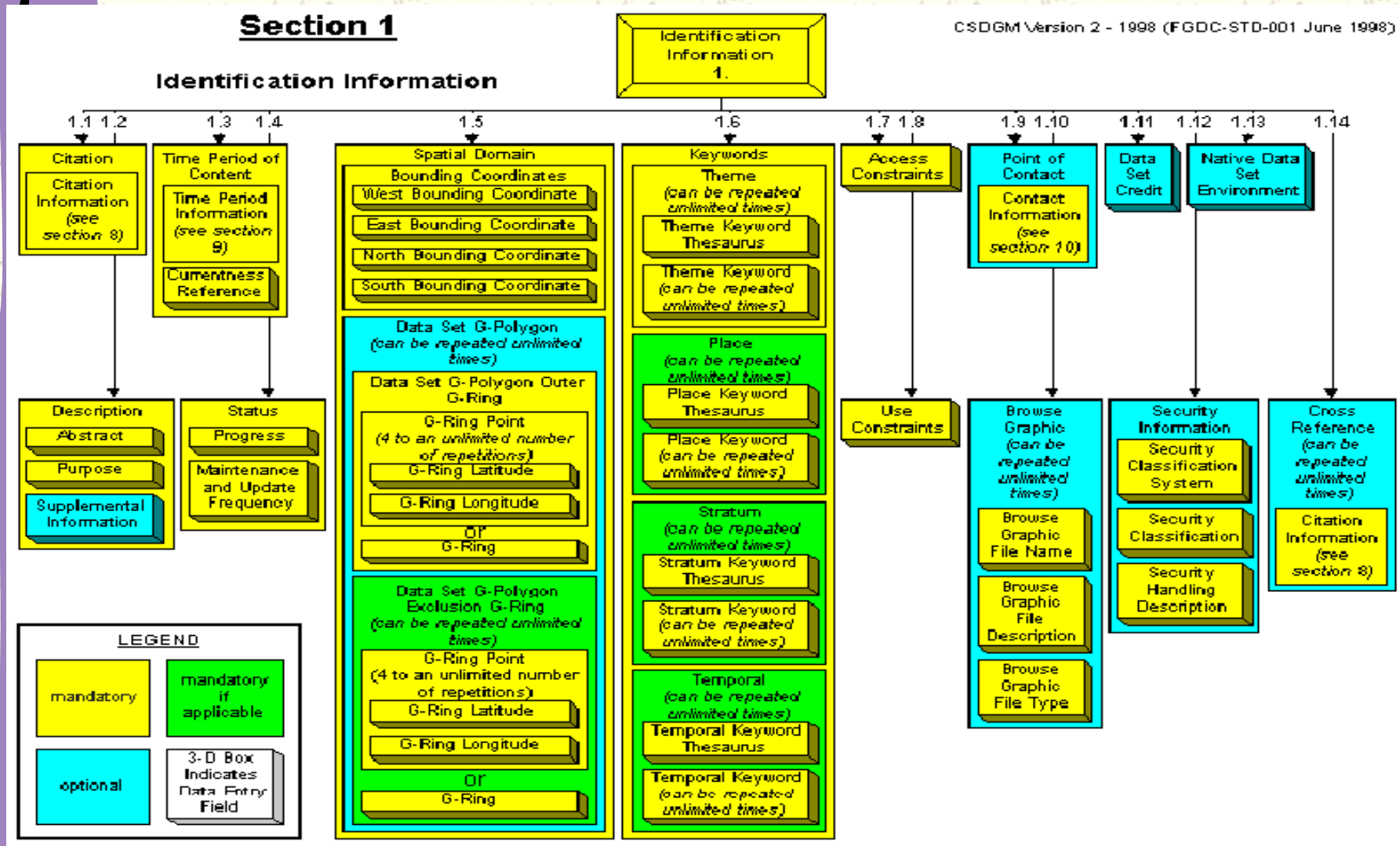
16

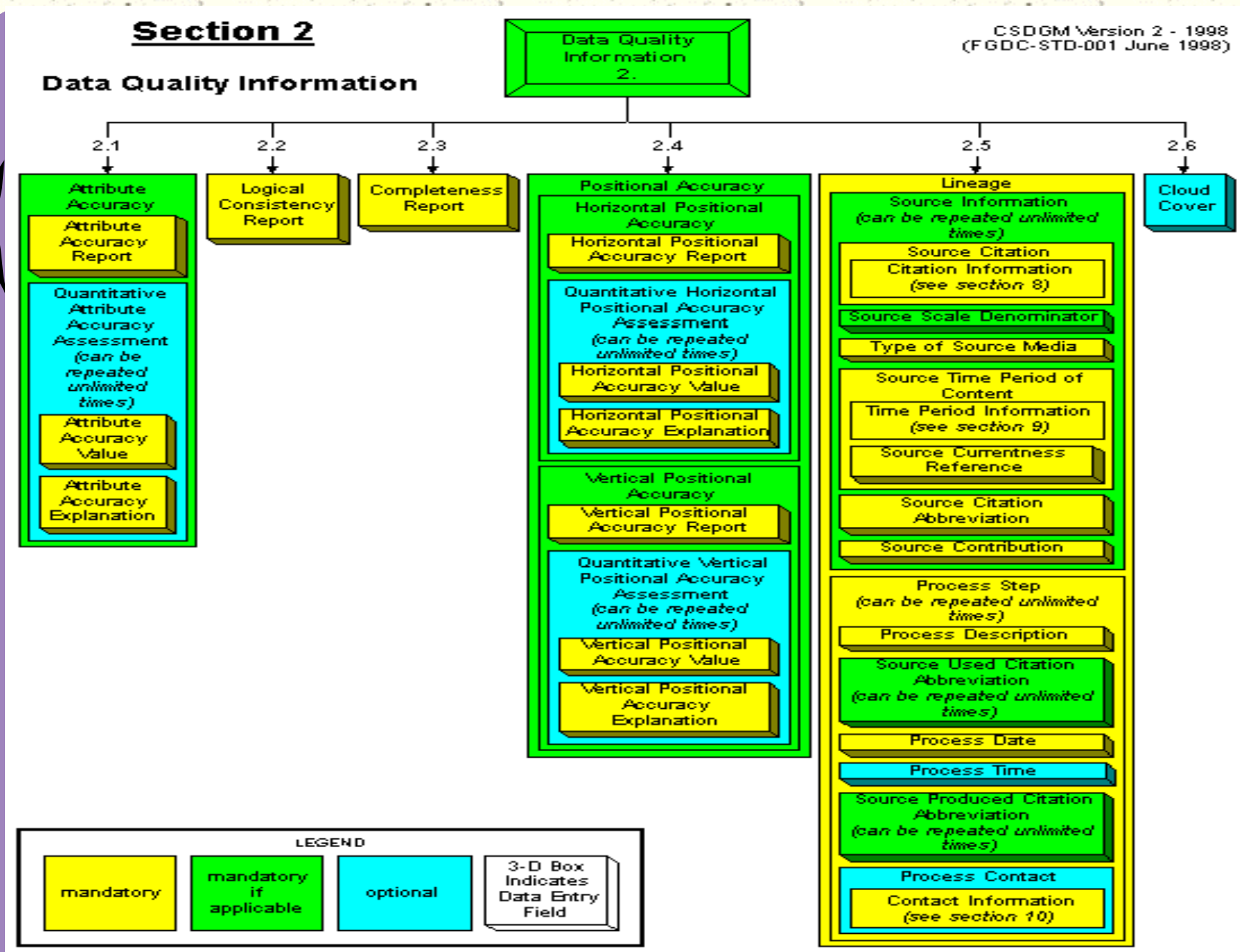
CSDGM

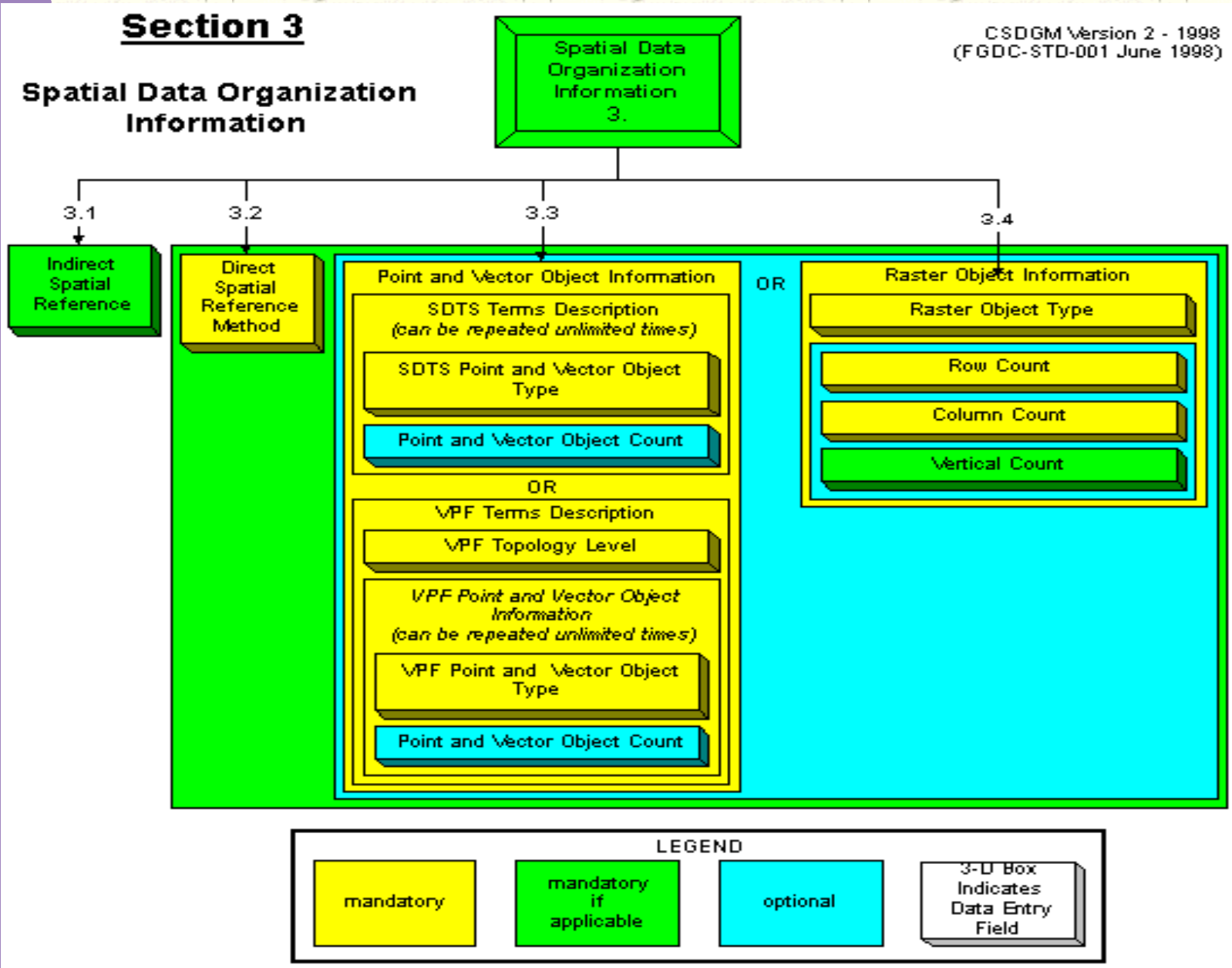
Production Rules



CSDGM

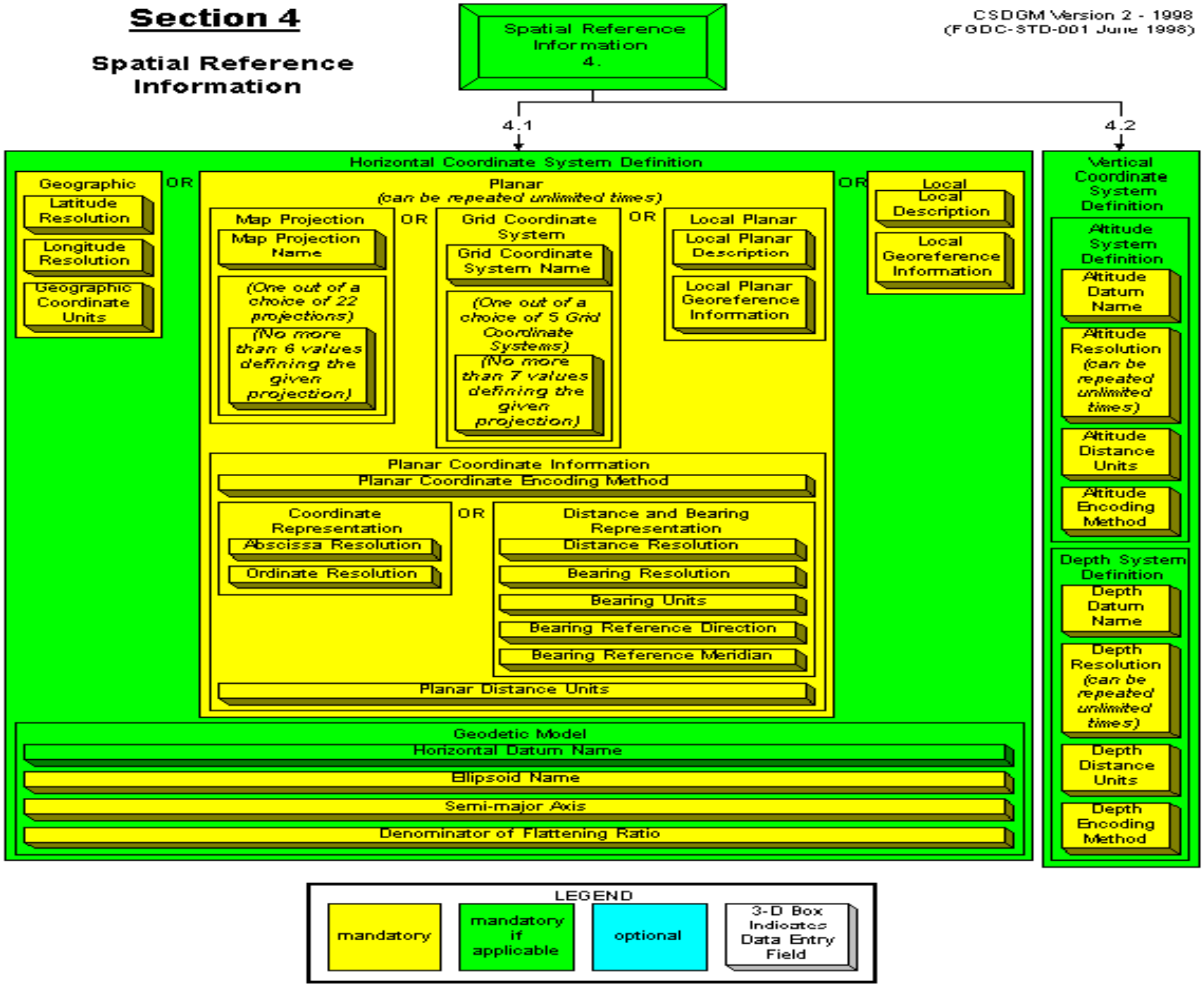

Section 1
Identification Information
CSDGM Version 2 - 1998 (FGDC-STD-001 June 1998)
Identification Information 1.
1.1 1.2
1.3 1.4
1.5
1.6
1.7 1.8
1.9 1.10
1.11
1.12
1.13
1.14
Citation
Citation Information (see section 8)
Time Period of Content
Time Period Information (see section 9)
Currentness Reference
Spatial Domain
Bounding Coordinates
West Bounding Coordinate
East Bounding Coordinate
North Bounding Coordinate
South Bounding Coordinate
Data Set G-Polygon (can be repeated unlimited times)
Data Set G-Polygon Outer G-Ring
G-Ring Point (4 to an unlimited number of repetitions)
G-Ring Latitude
G-Ring Longitude
Or
G-Ring
Data Set G-Polygon Exclusion G-Ring (can be repeated unlimited times)
G-Ring Point (4 to an unlimited number of repetitions)
G-Ring Latitude
G-Ring Longitude
Or
G-Ring
Keywords
Theme (can be repeated unlimited times)
Theme Keyword Thesaurus
Theme Keyword (can be repeated unlimited times)
Place (can be repeated unlimited times)
Place Keyword Thesaurus
Place Keyword (can be repeated unlimited times)
Stratum (can be repeated unlimited times)
Stratum Keyword Thesaurus
Stratum Keyword (can be repeated unlimited times)
Temporal (can be repeated unlimited times)
Temporal Keyword Thesaurus
Temporal Keyword (can be repeated unlimited times)
Access Constraints
Use Constraints
Point of Contact
Contact Information (see section 10)
Data Set Credit
Native Data Set Environment
Description
Abstract
Purpose
Supplemental Information
Status
Progress
Maintenance and Update Frequency
Browse Graphic (can be repeated unlimited times)
Browse Graphic File Name
Browse Graphic File Description
Browse Graphic File Type
Security Information
Security Classification System
Security Classification
Security Handling Description
Cross Reference (can be repeated unlimited times)
Citation Information (see section 8)
LEGEND
mandatory
mandatory if applicable
optional
3-D Box Indicates Data Entry Field

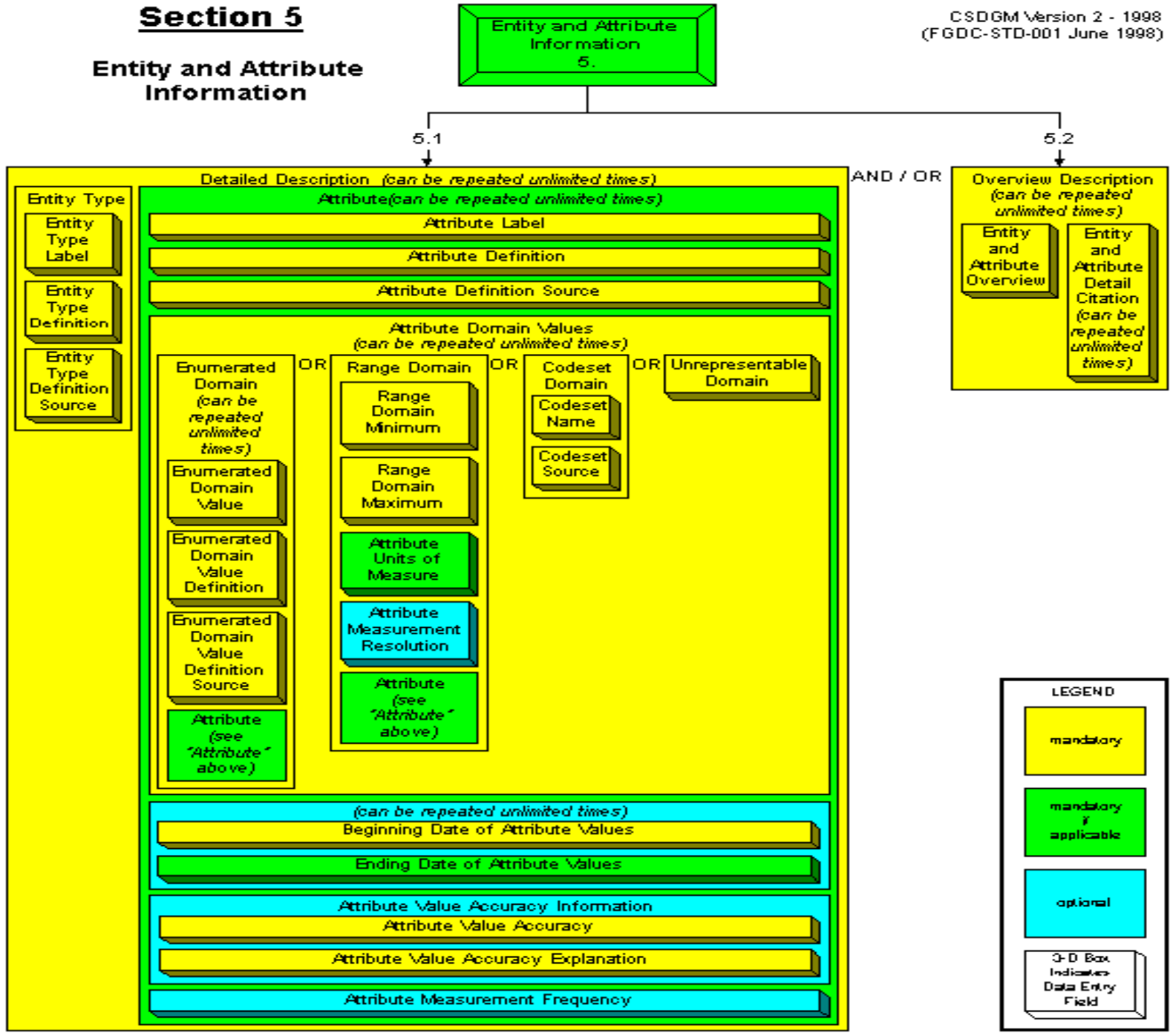

Section 2
Data Quality Information
Data Quality Information 2.
CSDGM Version 2 - 1998 (FGDC-STD-001 June 1998)
2.1
2.2
2.3
2.4
2.5
2.6
Attribute Accuracy
Attribute Accuracy Report
Quantitative Attribute Accuracy Assessment (can be repeated unlimited times)
Attribute Accuracy Value
Attribute Accuracy 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

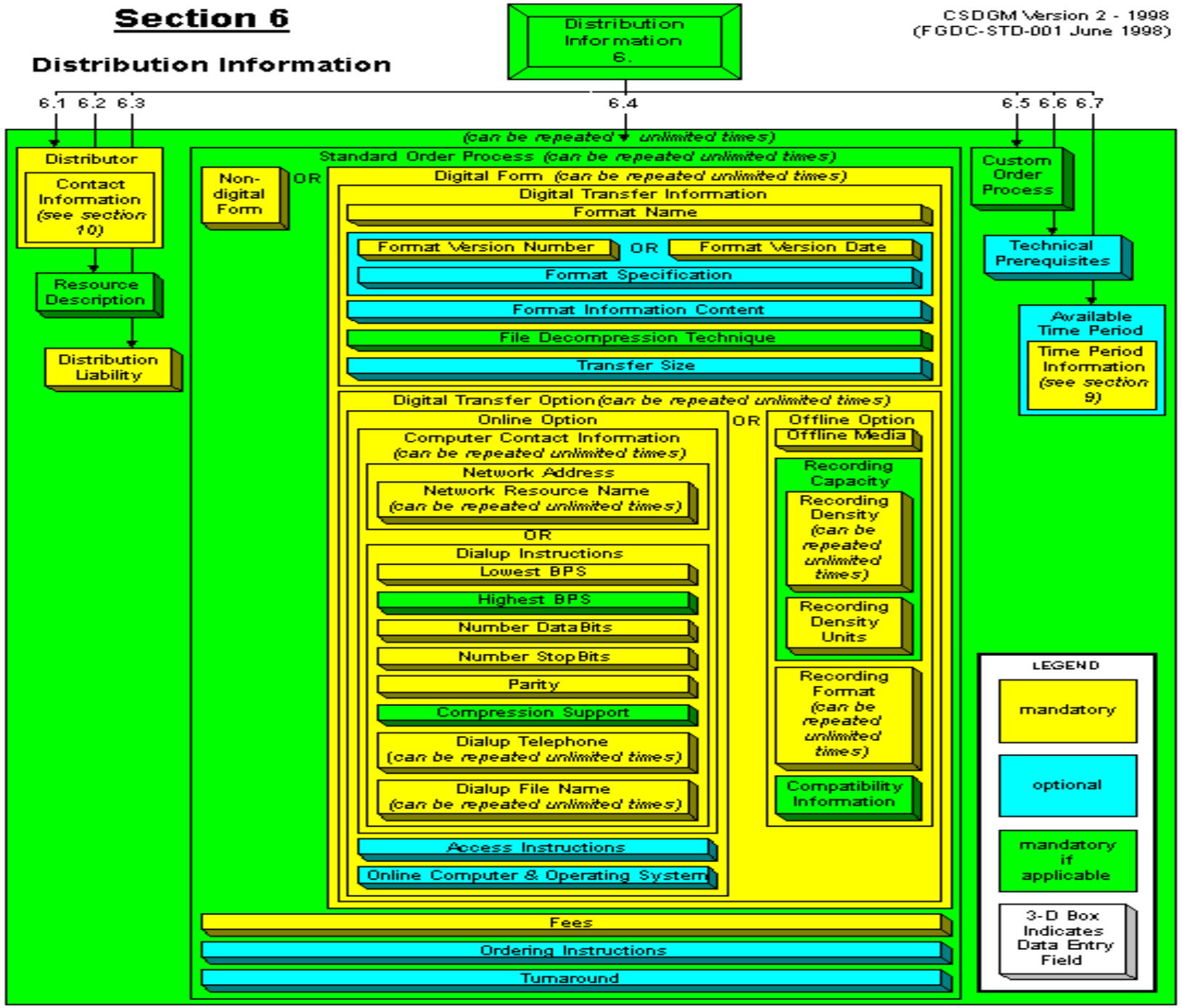

Section 3
Spatial Data Organization Information
Spatial Data Organization Information 3.
CSDGM Version 2 - 1998
(FGDC-STD-001 June 1998)
3.1
3.2
3.3
3.4
Indirect Spatial Reference
Direct Spatial Reference Method
Point and Vector Object Information
SDTS Terms Description
(can be repeated unlimited times)
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

Section 4

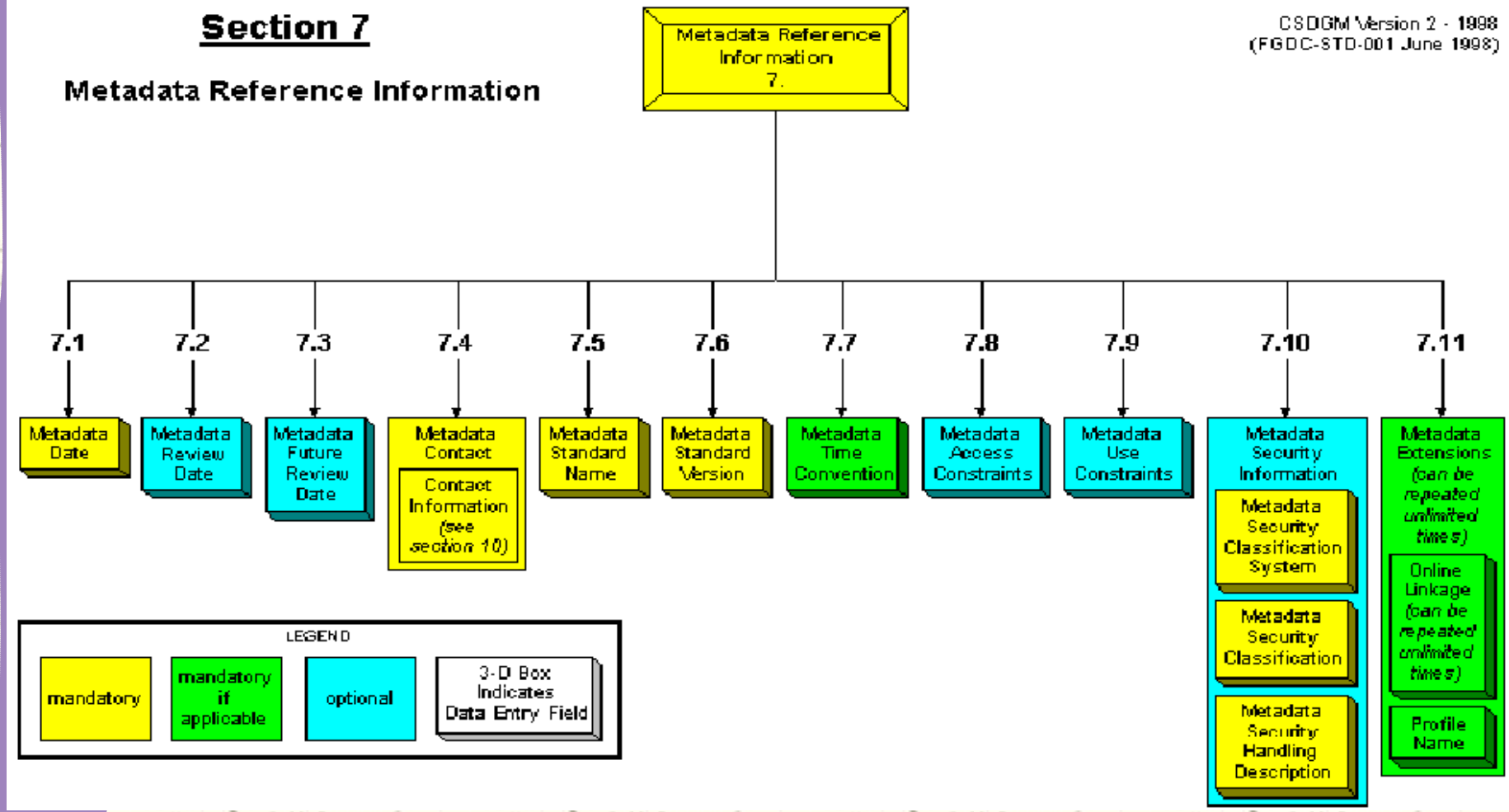
Spatial Reference Information




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

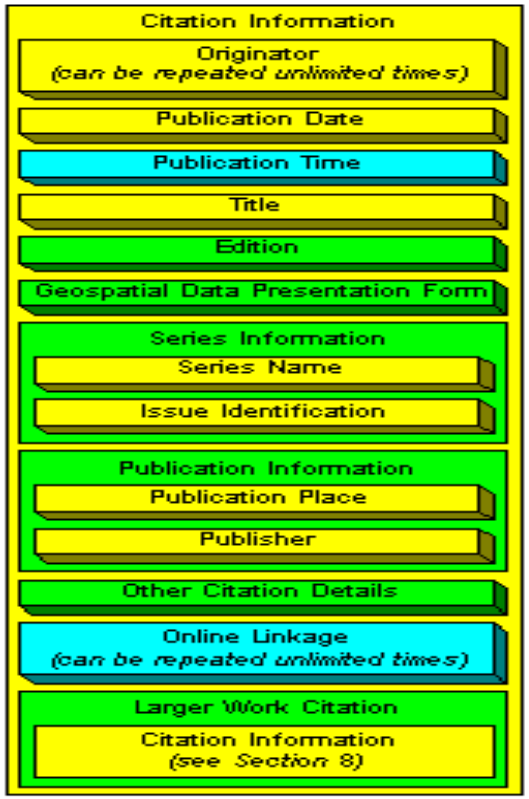

Section 6
Distribution Information
Distribution Information 6.
CSDGM Version 2 - 1998
(FGDC-STD-001 June 1998)
6.1 6.2 6.3
6.4
6.5 6.6 6.7
(can be repeated unlimited times)
Distributor
Contact Information (see section 10)
Resource Description
Distribution Liability
Standard Order Process (can be repeated unlimited times)
Non-digital Form
OR
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
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 Data Bits
Number Stop Bits
Parity
Compression Support
Dialup Telephone (can be repeated unlimited times)
Dialup File Name (can be repeated unlimited times)
Access Instructions
Online Computer & Operating System
OR
Offline Option
Offline Media
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

CSDGM


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

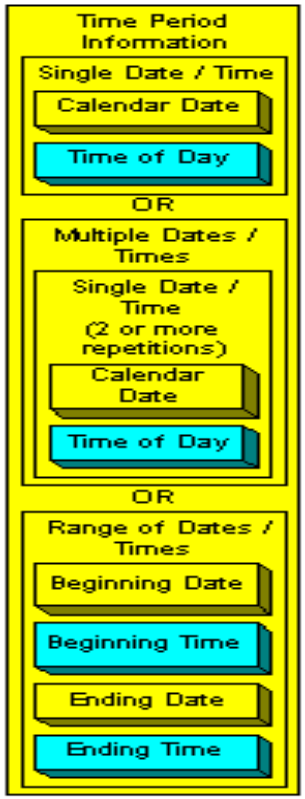
Section 8

Citation Information



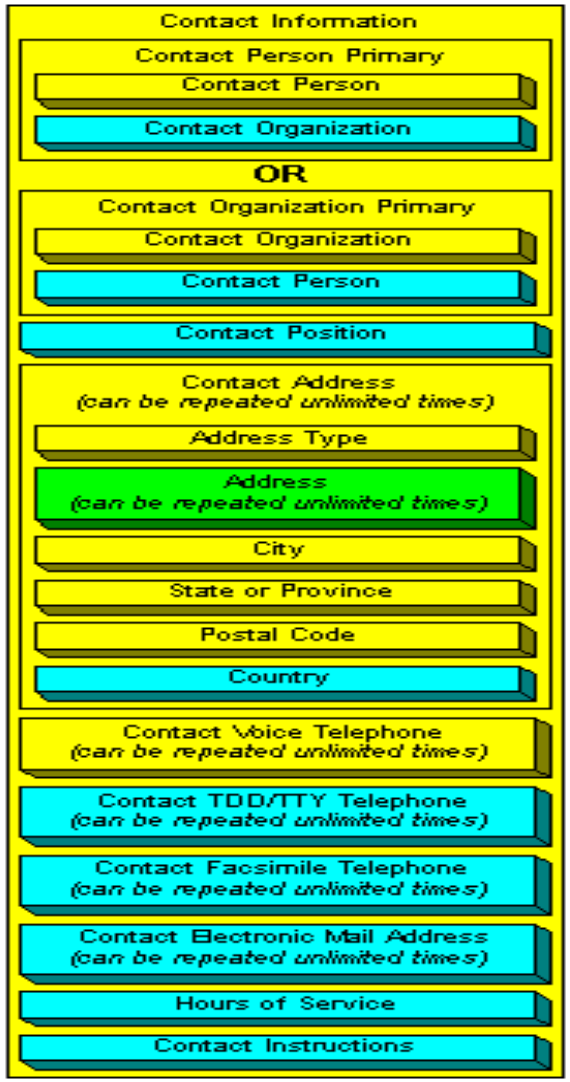
Section 9

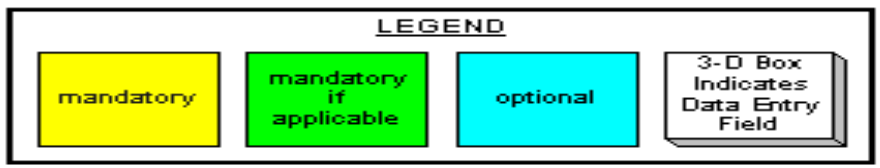
Time Period Information



Section 10

Contact Information





How Do We Get Started?

1. Select an appropriate metadata standard
2. Select an effective metadata creation tool
3. Find / review relevant example metadata records
4. Create the first record

Select a Metadata Tool

Metadata Collection

- Shareware
- GIS Internal
- Stand-alone Commercial

Metadata Validation

- pre-parser (format)
- parser (content)

Metadata Collection Tools

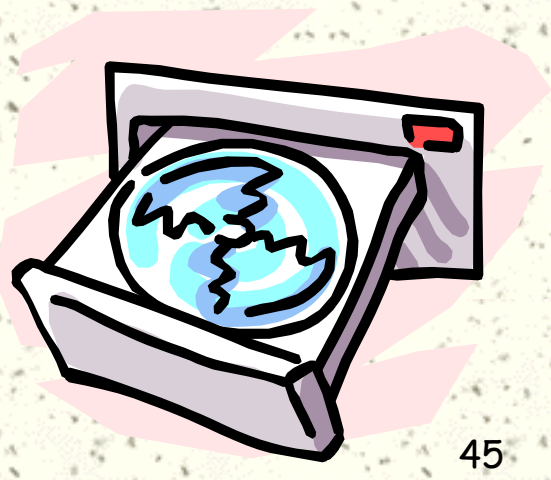
Shareware

- corpsmet - USACE
- metamaker - USGS/BRD
- tkme - USGS/FGDC
- metalite - USFWS
- M³Cat - ISO
- Enraemed - ISO (pending)

Metadata Collection Tools

GIS Internal

- ESRI ArcCatalog
- Intergraph SMMS for Geomedia



Metadata Collection Tools

Stand-alone Commercial

- SMMS
- Data Tracker

Metadata Collection Tools

Database / Spreadsheet

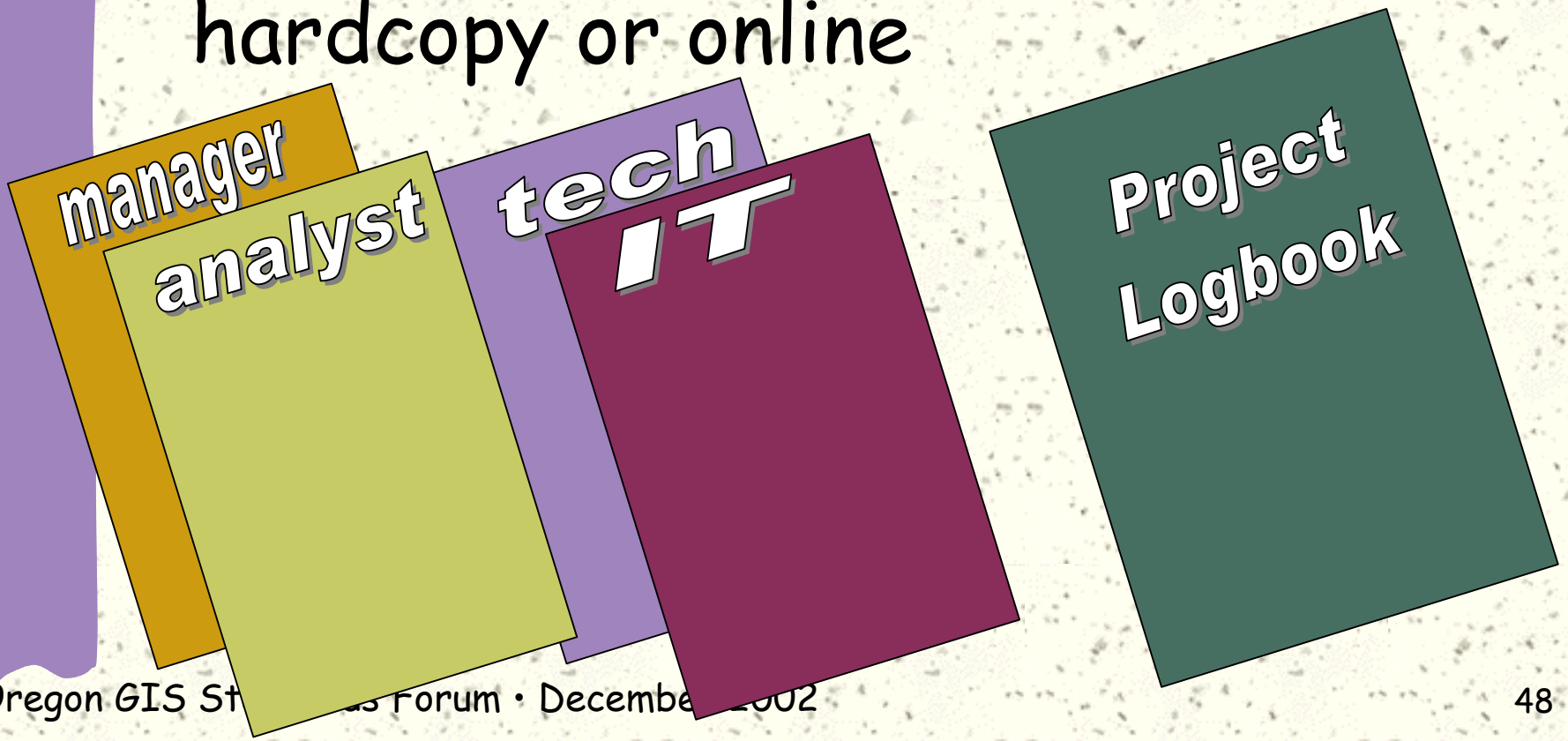
Access, Excel

METAFILE DATABASE	
Metafile_name	cadastral.met
Assoc. data file	cadastral.gis
Machine name/IP	130.39.128.555
Directory	d:\cherokee\cadastral.met
Last Update	04012000

Metadata Collection Tools

Forms

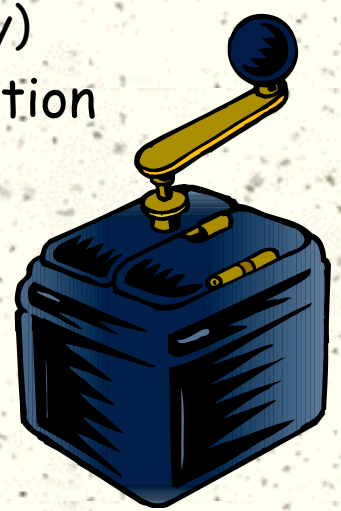
hardcopy or online



Metadata Validation Tools

- mp - Metadata Parser
  - Checks for CSDGM syntax
    - Element names
    - Mandatory elements
    - Element content (domains and logical consistency)
  - Embeds tags for NSDI Clearinghouse Distribution
- cns – Chew 'N Spit
  - Rectifies indentation
  - Inserts capitalization and underbars

FGDC 'compliant' metadata must pass mp!



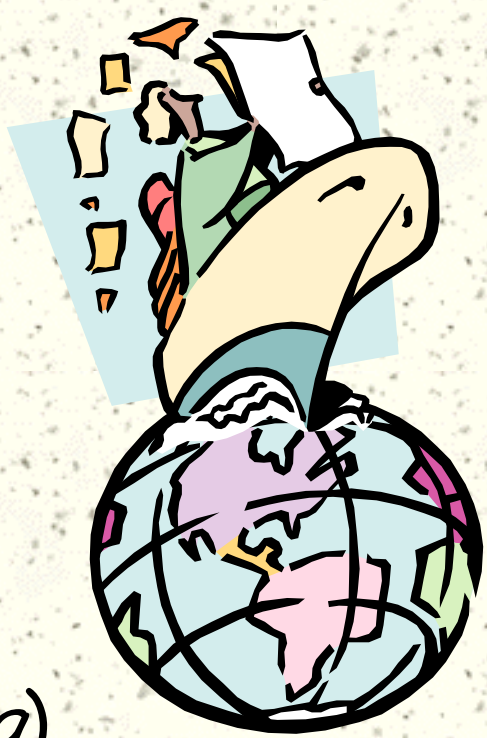
Review Metadata Examples

Clearinghouse

- NSDI
- Portland Metro
- Oregon Coast

FGDC Metadata Website

- Samples Library
- Online Resources (*pending*)



Thematic Agencies / Organizations

Review Metadata Examples

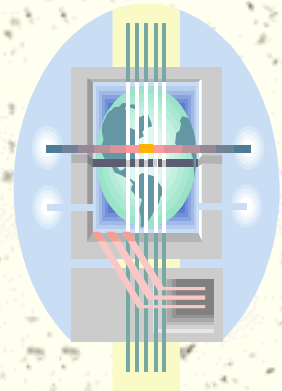
Metadata Format...*IMHO*

ASCII text is the most basic metadata transfer format

- XML emerging)

HTML is the easiest format to read/review

- can create using *mp*)

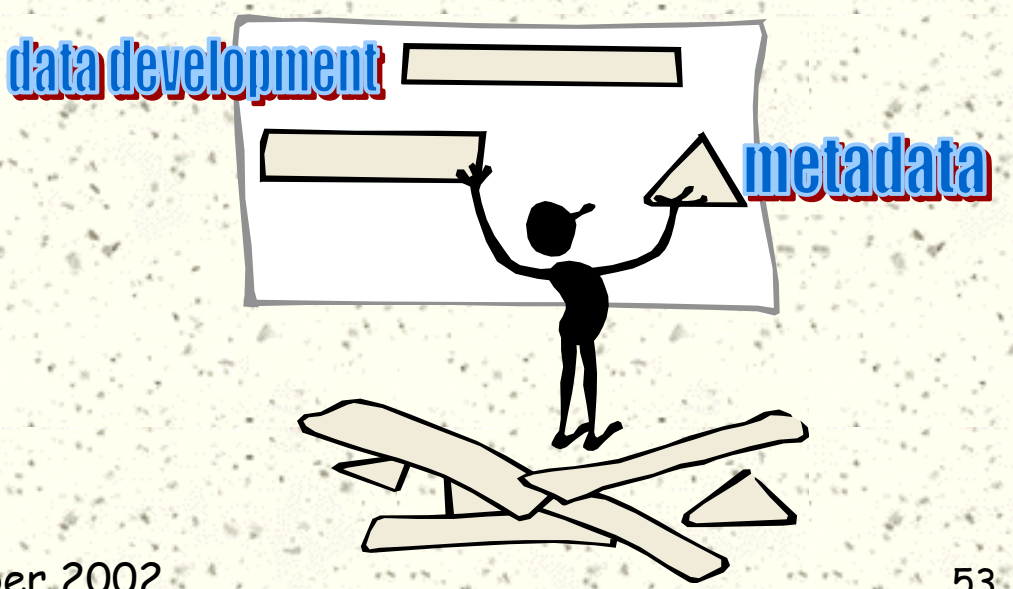


Create the First Record

- Interview all contributors
- Input information
- Use tool *help* features and available resources
- Check for grammar / spelling
- Review by all

Make Metadata Part of the Process

New tools enable us to better integrate metadata creation into the data development process



Make Metadata Part of the Process

If metadata were collected throughout the data process…

1. more accurate (no guessing)
2. more details
3. better decision-making.

Make Metadata Part of the Process

But what can I do to better incorporate metadata into the data development process?

Make Metadata Part of the Process

Step One: build administrative support

Make the business case

- preserve data investments
- limit data liability
- manage data resources
- find new data resources
- easier data transfer
- more efficient data distribution

Make Metadata Part of the Process

Step Two: build technical support

Emphasize individual benefits

- reduce workload..*in the long term*
- field fewer data inquiries
- document personal contributions

Make Metadata Part of the Process

Step Two: build technical support

Support your staff

- include in job descriptions & performance measures
- provide
  - staff support
  - tools
  - training

Make Metadata Part of the Process

Step Three: build templates

Organizational

- populate fixed fields
  - standardized language
  - distribution methods
  - standards used
- build source and contact libraries

Make Metadata Part of the Process

Step Three: build templates

Project

- abstract
- purpose
- keywords
- supplemental information
- geographic extent
- and more….

Make Metadata Part of the Process

Step Four: map fields to the workflow

Data Planning

Section 1: Identification Info

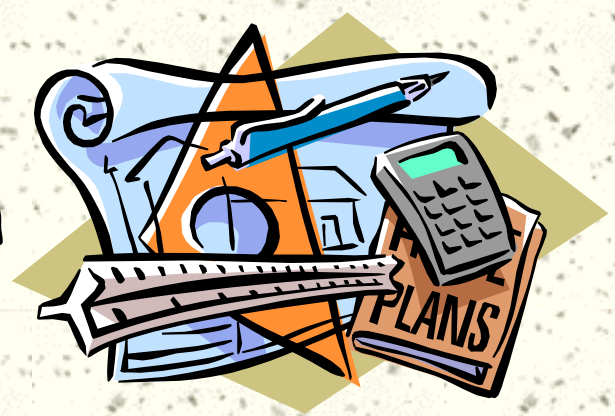
title, abstract, purpose, time period, keywords, extent

Section 2: Data Quality

sources, process methods

Section 3: Data Organization

point, vector, raster; indirect



Make Metadata Part of the Process

Step Four: map fields to the workflow

Data Planning

Section 4: Spatial Representation

coordinate system and parameters

Section 5: Entities and Attributes

entity/attribute labels and definitions

Section 6: Distribution (template)

Section 7: Metadata Reference

contact and date (template)

Make Metadata Part of the Process

Step Four: map fields to the workflow

Data Processing/Analysis

Section 2: Data Quality

process methods/contact, positional accuracy (RMS error), logical consistency (topological checks)

Section 5: Entities and Attributes

attribute values

Make Metadata Part of the Process

Step Five: distribute the effort

technicians - lineage

analyst - process methodology

scientists - accuracy assessments

IT managers - tools, collection, management

Make Metadata Part of the Process

Step Five: distribute the effort

Managers Can WRITE metadata

- data planning fields

and manage the effort

- metadata coordination
- metadata enforcement

Make Metadata Part of the Process

Step Six: policies and procedures

Establish Policies:

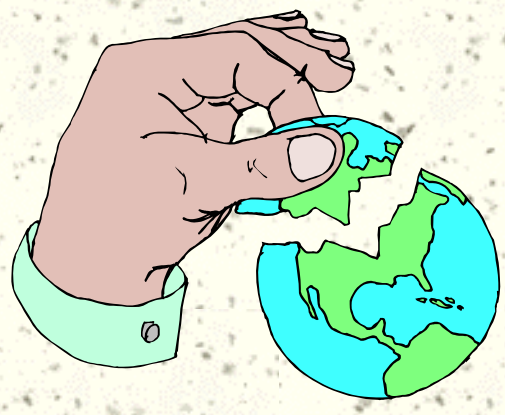
- mandate use of standards and templates
- develop boilerplate metadata deliverable language for data contractors
- require units to publish their metadata
- publish metadata SOP to document policies and procedures

Metadata as a Data Component

Metadata is that part of the data that:

- enables data discovery
- facilitates data import, display, management
- provides data consumer information
- instills data accountability
- limits data liability

Accurate metadata is fundamental to data quality



ISO 19115

- New International Metadata Standard
  - one component of larger geospatial standards effort
  - U.S. profile under development
    - in coordination with Canada and Mexico
  - will become ANSI standard

ISO 19115

- Fundamentals
  - heavy CSDGM influence
  - very robust (OO & UML)
  - highly normalized
  - new core fields
    - Language
    - Topic Category (hurray!)
  - metadata roll-up
  - tools already available!

topic categories

farming
biota
boundaries
climate / atmosphere
economy
elevation
environment
geo-science
health
imagery
military
inland waters
location
ocean
planning / cadastre
society
structure
transportations
utilities / communications

Contact Information

- Lynda Wayne
  Lwayne@GeoMaxim.com
  (828) 250.3801
- Cy Smith - OR State GIS Coordinator
  cy.smith@state.or.us
  (503) 378.6066
- FGDC
  www.fgdc.gov/metadata