

GOOD DATA MANAGEMENT & METADATA MAINTENANCE

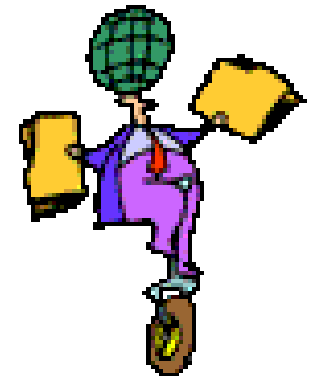
DC Metro Metadata Dissemination Seminars

Westat

The Major Uses of Metadata

2

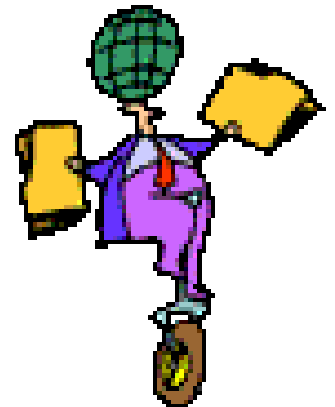
- To maintain an organization's internal **investment** in geospatial data



The Major Uses of Metadata

3

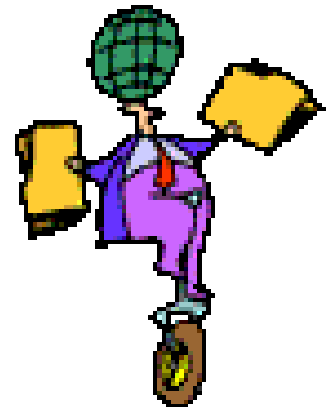
- To provide **information** about an organization's data holdings to data catalogues, clearinghouses, and brokerages



The Major Uses of Metadata

4

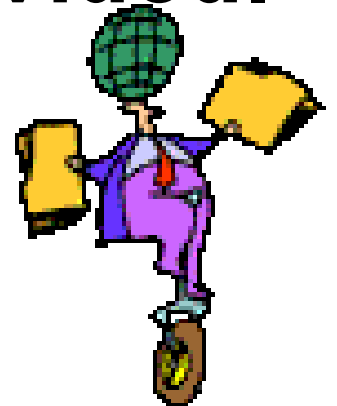
- **To provide information needed to process and **interpret** data to be received through a transfer from an external source**



The Major Uses of Metadata

5

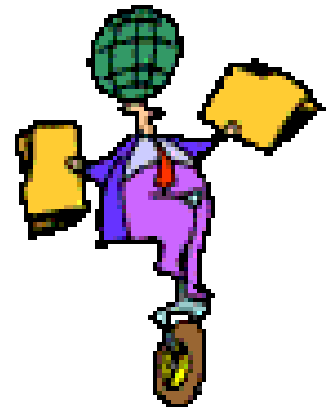
- To maintain **organized** data into several levels ranging from a simple listing of basic information about available data to detailed documentation about an individual data set



The Major Uses of Metadata

6

- To support the **creation** of an inventory of the data holdings of a state or local government agency



GOOD DATA MANAGMENT

How Do I Create Metadata?

8

- **Assemble information about the data set.**
- **Create a digital file containing the metadata, properly arranged**



How Do I Create Metadata?

9

- **Check the syntactical structure of the file**
 - ▣ **Modify the arrangement of information and repeat until the syntactical structure is correct**
- **Review the content of the metadata**
 - ▣ **Verifying that the information describes the subject data completely and correctly**

Parseable & Interoperable Metadata

10

Metadata must be both parseable, meaning machine-readable, and interoperable, meaning they work with software used in the Clearinghouse

Parseable Metadata

11

□ Parseable

- To ***parse*** information is to analyze it by disassembling it and recognizing its components

Parseable Metadata

12

- **Metadata that are parseable clearly separate the information associated with each element from that of other elements**

Parseable Metadata

13

- **The element values are not only separated from one another but are clearly related to the corresponding element names, and the element names are clearly related to each other as they are in the standard**

Parseable Metadata

14

- In practice this means that your metadata must be arranged in a hierarchy, just as the elements are in the standard, and they must use standard names for the elements as a way to identify the information contained in the element values**

Interoperable Metadata

15

- ☐ **To operate with software in the Clearinghouse, your metadata must be readable by that software**

Interoperable Metadata

16

- The data **must be parseable** and must identify the elements in the manner expected by the software

Interoperable Metadata

17

- **The FGDC Clearinghouse Working Group has decided that metadata should be exchanged in Standard Generalized Markup Language (SGML) conforming to a Document Type Declaration (DTD) developed by USGS in concert with FGDC**

What is in Good Metadata

18

- ☐ **Who**
- ☐ **What**
- ☐ **Where**
- ☐ **How**
- ☐ **When**
- ☐ **Why**

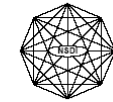


General Recommendations

19

- Use the 'Green Book'
(CSDGM Workbook)
- For DOQQ's and other large
data sets with many tiles...

Create a single metadata
record for the entire data set
if time and resources are
short, **BUT** try to create new,
more accurate metadata as
you build regional subsets



National Spatial Data Infrastructure

Content Standard for Digital Geospatial Metadata Workbook
(For use with FGDC-STD-001-1998)

Version 2.0

Federal Geographic Data Committee

May 1, 2000

Federal Geographic Data Committee

Department of Agriculture • Department of Commerce • Department of Defense • Department of Energy
Department of Health & Human Services • Department of Housing and Urban Development
Department of the Interior • Department of Justice • Department of State
Department of Transportation • Environmental Protection Agency
Federal Emergency Management Agency • Library of Congress
National Aeronautics and Space Administration • National Archives and Records Administration
National Science Foundation • Tennessee Valley Authority

Metadata Guide

20

1. **Identification**
2. **Data quality**
3. **Spatial data organization**
4. **Spatial reference**
5. **Entity and attributes**
6. **Distribution information**
7. **Metadata reference**



Identification

21

- **Originator (8.1)**
 - **Indicate the party responsible for the data set**

Identification

22

- **Publication Date (8.2)**
 - ▣ **The date that the data was published or otherwise made available.**

Remember format: YYYYMMDD

Identification

23

□ Title (8.4)

- Minimum - where, what, when,
- Best practice – who, why, resolution, filename, source

Identification

24

- **Online Linkage (8.10)**
 - ▣ **To provide access to a variety of data download, data clearinghouse, and web-mapping services**
 - ▣ **Use this field to fully represent your geospatial data access and distribution capabilities by providing complete URLs and necessary information to indicate the nature of the weblink**

Identification

25

□ Abstract (1.2.1)

- General content and features
- Data set form (GIS, CAD, image, Dbase)
- Geographic coverage (county/city name)
- Time period of content (begin and end date or single date)
- Special data characteristics or limitations

Identification

26

- **Supplemental Information**
 - **Place information that is not elsewhere covered**
 - **‘Front’ important information such as related studies, data set limitations, and notifications**

Identification

27

- **Time Period of Content (1.3)**
 - ▣ **The relevant date of the data content**
 - ▣ **Can be a single date, multiple dates, or a range of dates**



Identification

28

- **Currentness Reference (1.3.1)**
 - ▣ **The context for the time period of content**

Identification

29

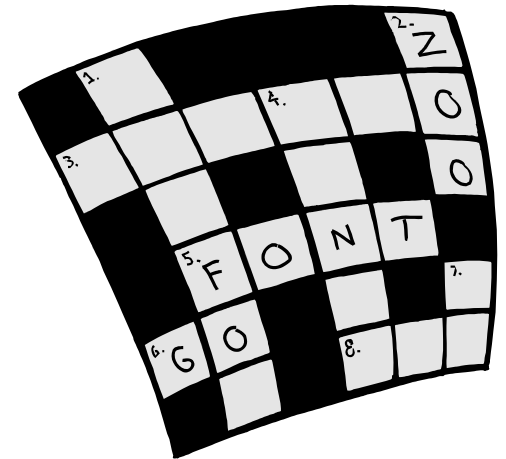
- **Progress (1.4.1)**
 - ▣ **The status of the data set, this field has a fixed domain of:**
 - **Complete**
 - **In work**
 - **Planned**



Identification

30

- **Theme Keyword (1.6.1.1)**
 - ▣ **Include broad and specific terms and use controlled vocabularies (thesauri) when possible**



Identification

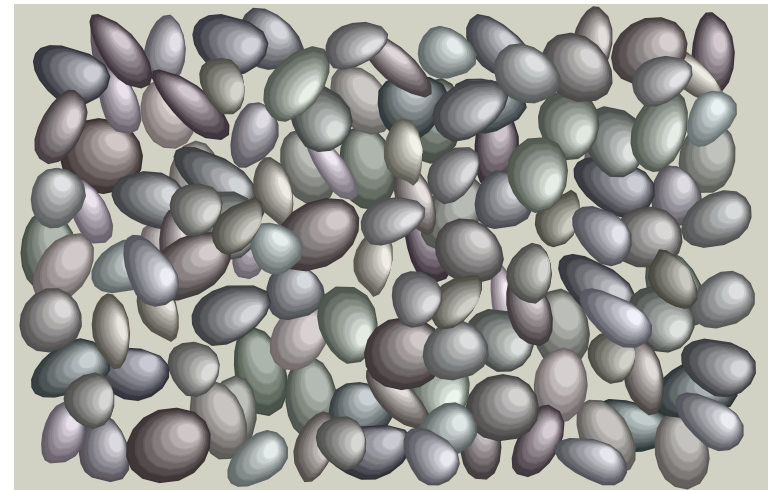
31

- **Place Keyword (1.6.2.1)**
 - **City or county name**
 - **State**
 - **State acronym**
 - **Regional descriptions and references**

Identification

32

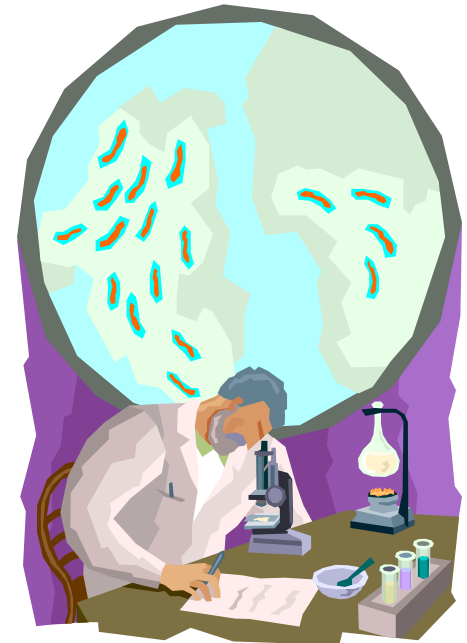
- **Stratum Keyword (1.6.3)**
 - ▣ **For use in atmospheric, geologic, and oceanographic data**



Identification

33

- **Temporal Keyword (1.6.4)**
 - ▣ **For use in scientific and historical data**



Identification

34

- **Access Constraints (1.7)**
 - **Any restrictions or legal prerequisites to accessing the actual data set**

Identification

35

- **Use Constraints (1.8)**
 - ▣ **Acknowledgement of the Originator when using the data set as a source**
 - ▣ **Sharing of data products developed using the source data set with the Originator**

Identification

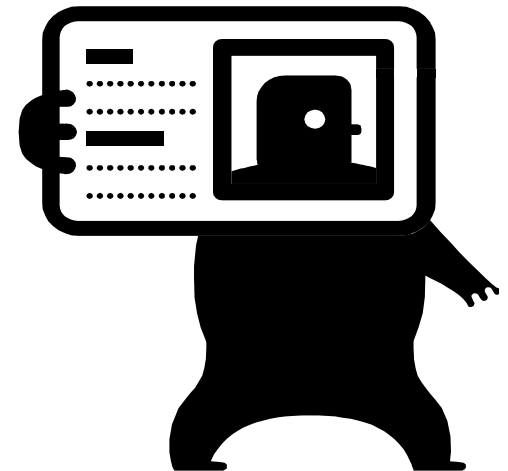
36

- **Use Constraints (1.8)**
 - ▣ **Data should not be used beyond the limits of the source scale**
 - ▣ **The data set is NOT a survey document and should not be utilized as such**

Identification

37

- **Point of Contact (1.9)**
 - ▣ **The individual or organization that is knowledgeable about the data set and should be contacted with questions**



Identification

38

- **Data Set Credit (1.11)**
 - ▣ **Identify others that should be recognized for their contributions to the data set**

Identification

39

- **Native Data Set Environment (1.13)**
 - ▣ **Software and version**
 - ▣ **Operating system and version**
 - ▣ **Platform**



Data Quality

40

- **Attribute Accuracy Report (2.1.1)**
 - ▣ **How sure are you?**



Data Quality

41

- **Logical Consistency Report (2.2)**
 - ▣ **Did you check for bad values and conditions?**

Data Quality

42

- **Completeness Report (2.3)**
 - ▣ **Is there anything I might expect to be in the data set that isn't?**

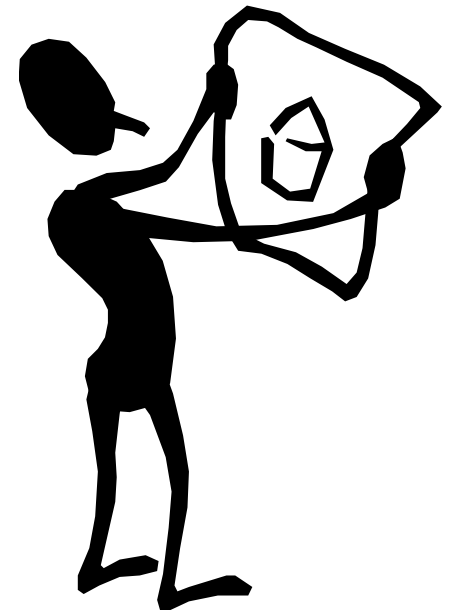


Data Quality

43

□ Positional Accuracy Report (2.4)

- ▣ How sure are you that the pine tree is where you say it is?



Data Quality

44

- **Process Step (2.5.2)**
 - ▣ **Stages of processing**
 - ▣ **Incorporation of sources**
 - ▣ **Project milestone**



Data Quality

45

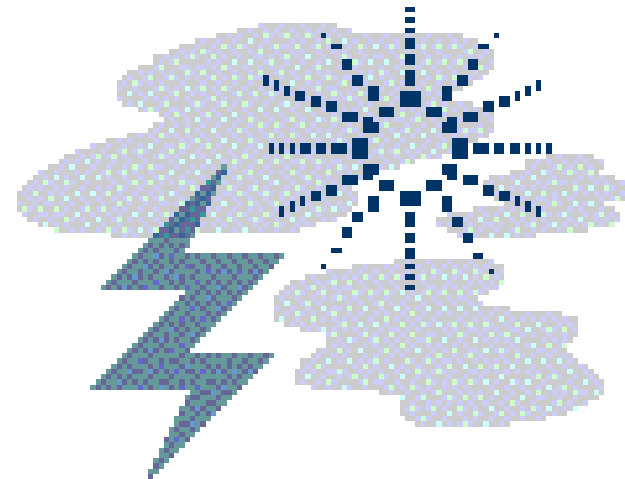
- **Process Contact (2.5.2.6)**
 - ▣ **The individual responsible for the data processing and ‘putting’ the data together**



Data Quality

46

- **Cloud Cover (2.6)**
 - ▣ **Leave blank for GIS and digital map files**
 - ▣ **Include values for imagery and photography**



Spatial Data Organization

47

- **Indirect Spatial Reference (3.1)**
 - ▣ **Geographic Names Index System (GNIS) place names**
 - ▣ **Public Land Survey System (PLSS) locations**
 - ▣ **Federal Information Processing System (FIPS) location codes**



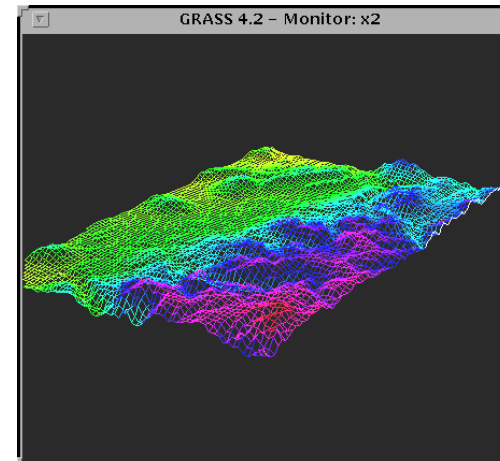
Spatial Data Organization

48

□ Direct Spatial Reference Method

(3.2)

- Vector
- Point
- Raster



Spatial Data Organization

49

- **SDTS Point and Vector Object Type (3.3.1.1)**
 - ▣ **For GIS files use 'Autocapture' feature of SMMS or ArcCatalog to populate**

Spatial Reference

50



□ Horizontal Coordinate System

Definition: (4.1)

- Geographic latitude / longitude
- Map projected
- Grid coordinate system
- Locally-defined oblique photography

Spatial Reference

51

- **Abcissa Resolution (4.1.2.4.2.1) / Ordinate Resolution (4.1.2.4.2.2)**
 - ▣ **The smallest distance that can exist between two points**
 - ▣ **The value is almost always the same for both the X axis (abscissa) and the Y axis (ordinate) but may differ for non-square pixels**

Spatial Reference

Abscissa/ Ordinate Resolution

52

□ Vector data

- ▣ This is commonly the ‘fuzzy tolerance’ or ‘clustering’ setting that establishes the minimum distance at which two points will NOT be automatically converged by the data collection device (digitizer, GPS, etc.)

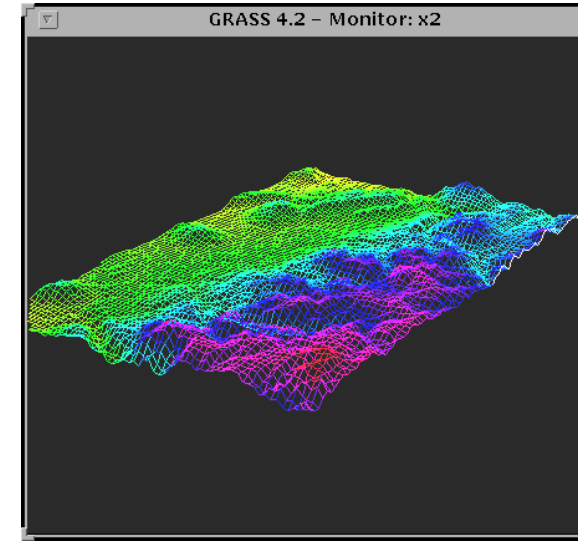
Spatial Reference

Abcissa/ Ordinate Resolution

53

□ Raster data

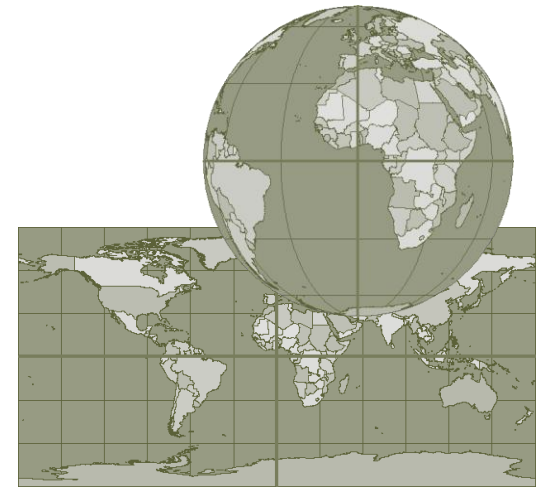
- ▣ The values normally represent the pixel size, e.g. for Thematic Mapper (TM) imagery, the value would be '30'



Spatial Reference

54

- **Planar Distance Units (4.1.2.4.4)**
 - ▣ **The units of measures for the Coordinate Representation (abscissa/ordinate resolution) or the Distance and Bearing Representation**



Entity and Attributes

55

- **Detailed Description (5.1)**
 - ▣ **Provide a detailed description if your database is not documented in another form such as a data dictionary or data specification manual**

Entity and Attributes

56

- **Attribute Domain Values (5.1.2.4)**
 - ▣ **Enumerated Domain a defined set of possible values, a picklist**
 - ▣ **Range Domain a sequence, series, or scale that has a defined maximum and minimum values - can be numeric or alphabetical**

Entity and Attributes

57

- **Attribute Domain Values (5.1.2.4)**
 - ▣ **Codeset Domain any published codeset**
 - ▣ **Unrepresentable Domain any value that is not prescribed**

Entity and Attributes

58

- **Overview Description (5.2)**
 - ▣ **Provide an overview description if your database is well-documented**
 - ▣ **Provide an overview description if your database is minimal and you can adequately describe in a short descriptive paragraph**

Distribution Information

59

- **Distributor Contact (6.1)**
 - ▣ **The individual or organization that distributes the data**



Distribution Information

60

- **Distribution Liability (6.3)**
 - ▣ **A statement of the liability assumed by the Distributor**
 - **Deny liability if the data are incorrect, incomplete, or misused**
 - **Limit third party distribution of the data set**

Metadata Reference

61

- **Metadata Date (7.1)**
 - ▣ **The date that the metadata is written or completed**



Metadata Reference

62

- **Metadata Contact (7.4)**
 - ▣ **The individual or organization that is responsible for the metadata for the data set**

Metadata Reference

63

- **Metadata Standard Name (7.5)**
- **Metadata Standard Version (7.6)**

Metadata Reference

64

- **Metadata Access Constraints (7.8)**
 - ▣ **Restrictions and legal prerequisites for accessing the metadata (not the data)**

Metadata Reference

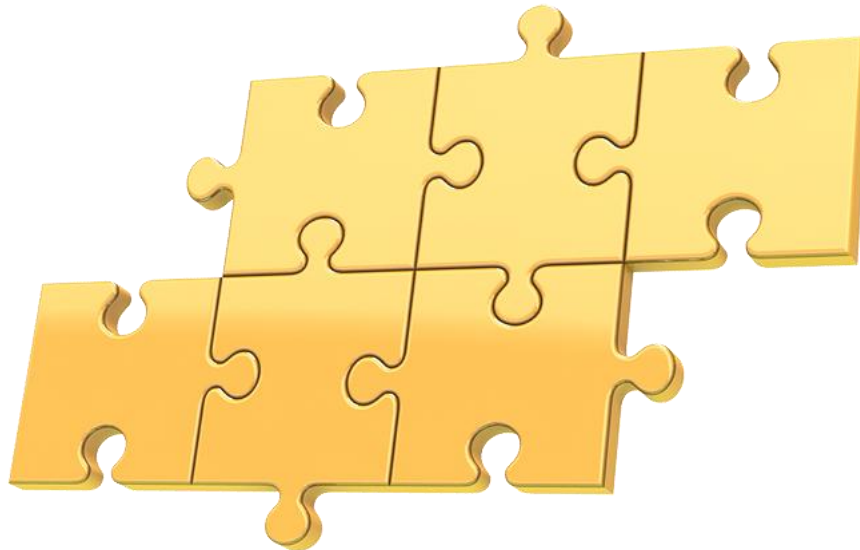
65

- **Metadata Use Constraints (7.9)**
 - ▣ **Restrictions and legal prerequisites for using the metadata (not the data) after access is granted**

Good Metadata

66

**Good Data has good
Metadata**



METADATA MAINTENANCE & TOOLS

Standards Defined

68

- **Webster's Definition**
 - ▣ **An acknowledged measure of comparison for quantitative or qualitative value; A criterion widely recognized as a model of authority or excellence**



Geospatial Standards

69

- **Types of standards**
 - **Management Standards**
 - **Data Standards**

Geospatial Standards

70

- **Geospatial Data Management**
 - **Metadata standards**
 - **Transfer standards**

Geospatial Standards

71

□ **Geospatial Data**

□ **Content**

- **Consistent structure for compatibility**

□ **Classification**

- **Forest categories**
- **Roads vs. Highways**

□ **Collection**

Content Standard

72

- **Defines what a thematic data layer should contain**
- **What are the fundamental elements of a cadastral layer?**

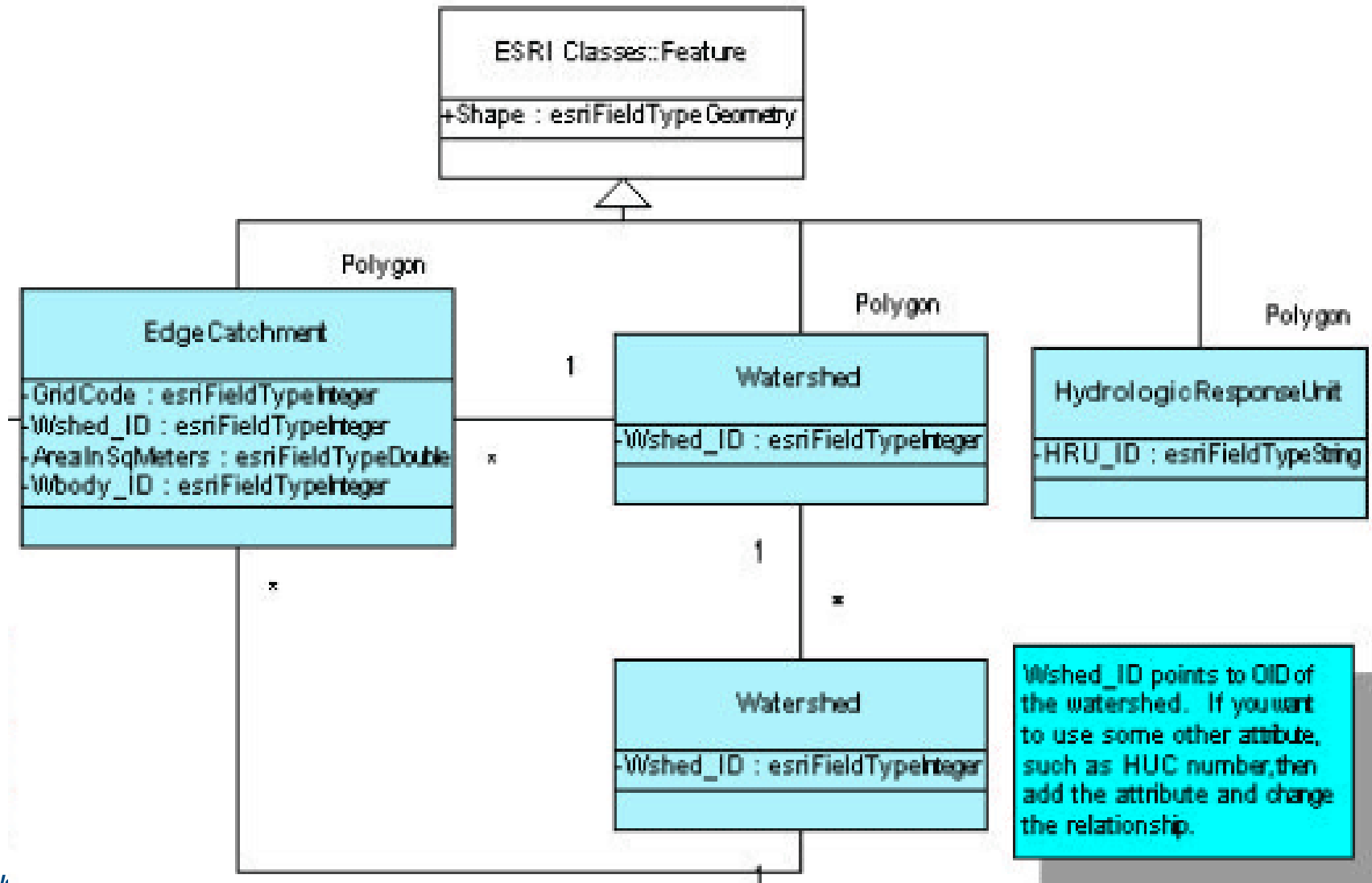
Content Standard

73

- **Semantic definition of a set of objects**
- **Organized and presented as a data model**
 - ▣ **Entity – relationship model**

Content Standard

74



Content Standard for Digital Geospatial Metadata (CSDGM)

75

- **The objectives of the standard are to provide a common set of terminology and definitions for the documentation of digital geospatial data**
- **Populate databases for consistent discovery and management**

CSDGM Formats

76

- **ASCII text document**
- **Hypertext Markup Language (HTML)**
- **Standard Generalized Markup Language (SGML)**

Description:

Abstract:

This is a point shapefile containing information about buoys.

Purpose:

This data is being used to teach ArcCatalog's FGDC metadata editor.

Time_Period_of_Content:

Time_Period_Information:

Databuoy

Metadata:

- [Identification Information](#)
- [Spatial Data Organization Information](#)
- [Spatial Reference Information](#)
- [Entity and Attribute Information](#)
- [Distribution Information](#)
- [Metadata Reference Information](#)

Identification_Information:

Citation:

Citation_Information:

07DA007B2DA6}</MetalID>

<CreaDate>20041102</CreaDate>

<CreaTime>17190800</CreaTime>

<SyncOnce>FALSE</SyncOnce>

<SyncDate>20041103</SyncDate>

<SyncTime>16132900</SyncTime>

<ModDate>20041103</ModDate>

<ModTime>16132900</ModTime>

</Esri>

- <idinfo>

<native Sync="TRUE">Microsoft Windows
2000 Version 5.0 (Build 2195) Service
Pack 3; ESRI ArcCatalog
9.0.0.535</native>

- <descript>

<langdata Sync="TRUE">en</langdata>

<abstract>This is a point shapefile

Metadata Creation & Management Software

77

- **Computer software to manage and implement metadata standards for an organization are wide-ranging**

When Choosing Software

78

- **Do our current systems already have a metadata capability?**
- **What need is there for integrating with other systems?**
- **What do our clients/customers need?**

When Choosing Software

79

- ☐ **What is the software cost?**
- ☐ **Is the software a flat file or database system?**
- ☐ **What file type is the output?**

Metadata Creation & Management Software

80

- ❑ **Various software packages specialize in metadata management or manipulation while others are a segment of a GIS software suite**

How do I Determine the Best Metadata Tool?

81

- **There are a variety of free and commercial software tools available to support metadata development**
- **These tools offer a range of features and capabilities**



How do I Determine the Best Metadata Tool?

82

- **When evaluating the right tool consider the following:**
 - ▣ **Does your GIS software provide an internal data/metadata management utility that supports the auto-capture of data set properties (extent, projection, attribute labels, etc.)?**

How do I Determine the Best Metadata Tool?

83

- **When evaluating the right tool consider the following:**
 - ▣ **Do you need a tool that supports one or more CSDGM Profiles or Extensions?**

How do I Determine the Best Metadata Tool?

84

- **When evaluating the right tool consider the following:**
 - ▣ **Do you need to document data resource beyond those managed within your GIS?**

How do I Determine the Best Metadata Tool?

85

- **When evaluating the right tool consider the following:**
 - ▣ **Do you need a tool that is easily distributed to partners?**

How do I Determine the Best Metadata Tool?

86

- **When evaluating the right tool consider what features are important:**
 - ▣ **Auto-capture of information**
 - ▣ **Storing metadata with data**
 - ▣ **Creation and use of templates**
 - ▣ **Easy to use interface**
 - ▣ **Robust help systems and tutorials**

Metadata Creation & Management Software

87

- The list provided here is not an endorsement of any given software, but a list gathered from personal use and other trainer recommendations**

FGDC Metadata Tools?

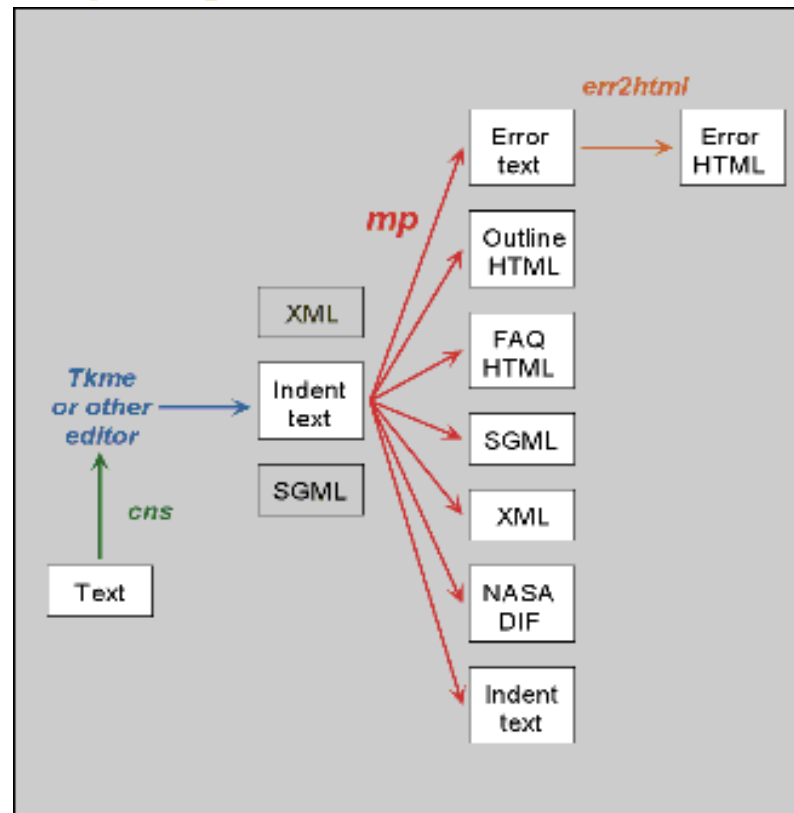
88

The FGDC does not promote the use of any specific metadata tool. It does however, test compliance to the CSDGM using the Metadata Parser (mp) utility developed by Peter Schweitzer of the USGS Geology Discipline.

Metadata Parcer

89

- <http://geology.usgs.gov/tools/metadata/tools/doc/mp.html>



Creation and Editing Tools

Commercial tools:

90

□ **ArcCatalog**

<http://www.esri.com/products/index.html>

□ **Intergraph**

▣ **GeoMedia Catalog**

▣ **SMMS**

<http://www.intergraph.com/customerstories/CIESIN.aspx>

Creation and Editing Tools

Freeware/Shareware tools:

91

☐ **tkme**

<http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html>

☐ **xtme**

<http://geology.usgs.gov/tools/metadata/tools/doc/xtme.html>

☐ **MetaScribe**

<http://www.csc.noaa.gov/metadata/metascibe/>

Creation and Editing Tools

Freeware/Shareware tools:

92

- **MERMAid**

<http://www.ncddc.noaa.gov/activities/mermaid/>

- **Metavist 2005**

<http://ncrs.fs.fed.us/pubs/viewpub.asp?key=2737>

- **EPA Metadata Editor (EME)**

<http://www.epa.gov/geospatial/eme.html>

CSDGM Compliance Tools

Freeware/Shareware tools:

93

- **Metadata Parser**

<http://geology.usgs.gov/tools/metadata/tools/doc/mp.html>

- **Chew-N-Spit**

<http://geology.usgs.gov/tools/metadata/tools/doc/cns.html>

- **Geospatial Metadata Validation Service**

<http://geo-nsdi.er.usgs.gov/validation/>

What Tools are Available to Check the Accuracy of Metadata?

94

- ☐ **No tool can check the accuracy of metadata**
- ☐ **Human review is required**
- ☐ **Develop SOP's for flagging edits to data that may require metadata update**

Best Practices

95

- **What are they?**
 - ▣ **Represent effective and efficient methods for accomplishing common tasks**
 - ▣ **Similar to standards**
- **Not applicable to every situation**

Best Practices with Data

96

- ❑ **Enforce a common Spatial Reference System**
 - ❑ **e.g. State Plane Zone and Datum**
- ❑ **Consistent use of persistent ID's for objects**
- ❑ **Minimum Mapping Unit**
- ❑ **Change Report**

Best Practices with Metadata

97

- ❑ **Evaluate how others are operating**
 - ❑ **Identify advanced metadata users**
- ❑ **Require metadata with data collection and delivery**
- ❑ **Designate a “Metadata Manager”**
- ❑ **Develop a capture, maintenance and review process**

Best Practices with Metadata

98

- **Automate collection and maintenance**
 - **Templates and Forms**
- **Standardize consistent content**
 - **e.g. contact and distribution information**
- **Utilize change report to prioritize metadata review**
 - **Can detect if metadata record “date” corresponds with data layer change**

Thank You!

99

U.S. Geologic Survey

CAP 1 Grant for funding this training