

Describing data sets

IST400/600

Jian Qin

What's metadata?

- "...structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information".

Source: NISO (2004). *Understanding Metadata*. Bethesda, MD: NISO Press. <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

A few important terms

- **Metadata** are structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities.
- A **metadata schema** provides a formal structure designed to identify the knowledge structure of a given discipline and to link that structure to the information of the discipline through the creation of an information system that will assist the identification, discovery, and use of information within that discipline.
- **Interoperability** is the ability of two or more systems or components to exchange information and use the exchanged information without special effort on either system.

Source: CC:DA Task Force on Metadata: *Final Report* (CC:DA/TF/Metadata/5), June 16, 2000
<http://www.libraries.psu.edu/las/jca/ccda/TF-meta6.html>

Metadata

3

Metadata record

- A metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource.
 - > Information *about*
 - Who created the data resource
 - What there is in the data file(s)
 - When the data was captured/collected
 - Where the data was captured/collected
 - Why the data was captured/collected
 - How the data was captured/collected

Metadata

4

Data or metadata?

○ Data

- > Abstract: text
- > Purpose: text
- > Spatial domain:
{W, E} lon, {N, S} lat
- > Network resource: text

○ Metadata

- > Abstract: data set narrative summary
- > Purpose: why data set was developed
- > Spatial domain: geographic areal domain of data set
- > Network resource: data set URL

Metadata

5

Kinds of metadata (1)

○ Structural

- > Organization
 - Simple types
 - Structured types
- > Packaging
 - Formats

○ Source

- > Origin
 - Person
 - Organization
- > Lineage
 - Processing history
- > Quality
 - Precision
 - Accuracy

Metadata

6

Kinds of metadata (2)

- Annotation
 - › Abstracts
 - › References
 - › User feedback
- Availability
 - › Restrictions
 - Proprietary
 - Security
 - › Access
 - Handle
 - Online: e.g. URL
 - Offline: e.g. ISBN
 - Point-of-contact

Metadata

7

Some metadata idioms

- Control document reference
 - › Format description
 - › Compilation protocol
 - › Accuracy standard
- Point of contact
 - › Name
 - › Address
- Date
 - › Single
 - › Multiple
 - › Range
- Geolocation
 - › Lat, lon, datum
 - › X, y, {projection}

Metadata

8

Hierarchical metadata

- Most metadata schemes are hierarchical
 - Parent: class-level
 - Inherited by all members of class
 - e.g. map series
 - Child: item-level
 - Supplements or overrides parent metadata
 - e.g. individual map

Metadata

9

Metadata standards

Content Standard
for Digital
Geospatial
Metadata
(CSDGM)

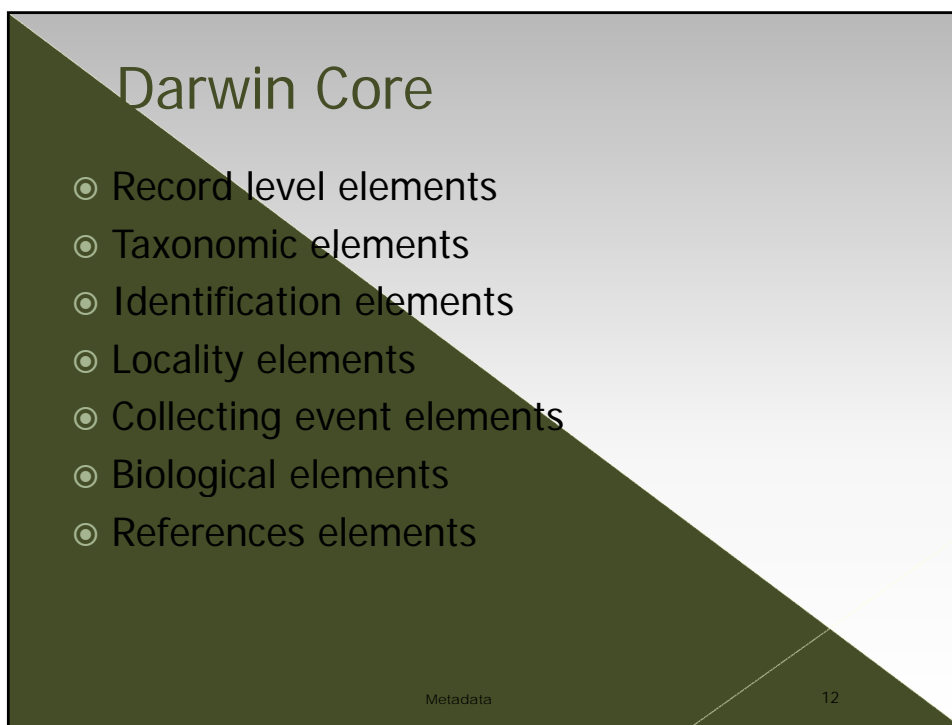
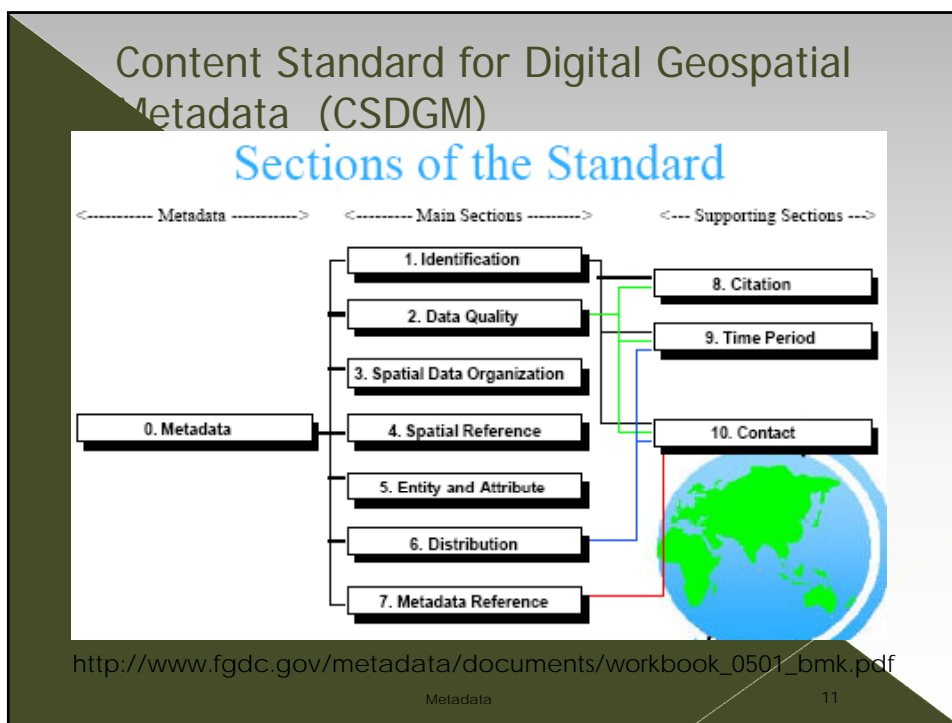
Darwin Core (DwC)

Dublin Core
Metadata Element
Set (DC)

Climate and
Forecast (CF)
Metadata
Convention

Metadata

10



The Dublin Core Metadata Element Set

- > Title
- > Subject
- > Description
- > Creator
- > Publisher
- > Contributor
- > Date
- > Type
- > Format
- > Identifier
- > Source
- > Language
- > Relation
- > Coverage
- > Rights

Metadata

13

Climate and Forecast (CF) Metadata Convention

- CF metadata conventions are designed to promote the processing and sharing of data stored in files created with the netCDF API
 - > netCDF files and components
 - > Description of data (units, long name, standard name, etc.)
 - > Coordinate types (lon, lat, vertical, time)
 - > Coordinate systems (independent lat, lon, vertical, and time axes, 2-d lat, lon, coordinate variables, etc.)
 - > Labels and alternative coordinates
 - > Data representative of cells (cell boundaries, measures, methods, etc.)
 - > Reduction of dataset size (packed data, compressed data, etc.)

<http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.1/cf-conventions.pdf>

Metadata

14

A case study

- Datasets for a major outbreak of severe weather and tornado:
http://data.eol.ucar.edu/codiac/projs?COMET_CASE_044
- > Description levels
 - The overall description
 - Individual datasets
- > Datasets
 - What elements are used to describe the datasets?
 - How would you reorganize the dataset information?
 - Search and browse scenarios