

## 4. METADATA STANDARDS

Metadata, or data about data, describe the content, quality, condition, and other characteristics of data. For example, metadata for vertical profiles of ocean properties obtained from hydrographic casts might describe, minimally, the locations and times of the casts, the inclusive depths, the variables measured, the location of the data, and the name of the person to contact to request access to the data. In general, metadata include thematic, semantic and syntactic descriptors of the data they reference.

- Thematic metadata describe the context of the study that produced the data. Such descriptors can include, for example, principal investigator, species association and study hypothesis.
- Semantic metadata describe contextual information about the data. Candidate descriptors are measurement type, measurement device, units of measurement, calibration information, *etc.*
- Syntactic descriptors define the way the data are packaged, *e.g.*, file size, file format, storage mechanism and location.

Metadata are coded using standard transcriptions developed over the years for different purposes. For example, metadata in NPEM are described using a common set of terminology and definitions provided by the FGDC metadata standard. The data themselves are not part of the metadatabase and continue to reside with their owner. Each metadata record provides a dynamic link to the data or to the contributor. The metadata standard determines what thematic, semantic and syntactic descriptors are catalogued and how they are presented. There are a number of metadata standards, *e.g.*, Directory Interchange Format (DIF), Ecological Metadata Language (EML), Dublin Core (DC), and FGDC.

The standard coding provides a basis for search methods. When seeking a certain kind of data, a scientist or manager uses a search procedure to examine a broad body of metadata. The search procedure examines all metadata records,

eliminating those that do not satisfy the search criteria. For example, when searching for vertical profiles of ocean properties obtained from hydrographic casts, a manager might specify the locations and times of the casts, the inclusive depths and the variables measured. The successful search will reiterate these parameters and tell the searcher the location of the data, and how those data can be obtained.

Committing to a metadata standard is not a trivial decision. In this section, we review the major metadata standards so that potential Federation participants can evaluate the benefits and disadvantages of adopting alternative metadata standards.

### 4.1. Directory Interchange Format

Directory Interchange Format (DIF, <http://gcmd.gsfc.nasa.gov/difguide/difman.html>) is a *de facto* standard used to create directory entries that describe a group of data. A short history of DIF can be found at <http://gcmd.nasa.gov/User/difguide/whatisadif.html>.

DIF is used to create directory entries that describe a group of data. DIF consists of a collection of fields that detail specific information about the data. Eight fields are required in DIF; the others expand upon and clarify the information. Some of the fields are text fields; others require the use of controlled keywords (sometimes known as “valids”). The eight required fields are:

- Entry\_ID,
- Entry\_Title,
- Science\_Keywords,
- ISO\_Topic\_Category,
- Data\_Center,
- Summary,
- Metadata\_Name, and
- Metadata\_Version.

As with most metadata standards, DIF allows users of data to understand the contents of a data set, and contains those fields that are necessary for users to decide whether a particular data set would be useful for their needs.

The most well-known and widely used source for DIF metadata format is that employed by the Global Change Master Directory (GCMD, <http://gcmd.nasa.gov/>).

## 4.2. Ecological Metadata Language

Ecological Metadata Language (EML) is a metadata specification developed by the ecology discipline and supported by The Knowledge Network for Biocomplexity (<http://knb.ecoinformatics.org/home.html>). It is based on prior work done by the Ecological Society of America and associated efforts (Michener *et al.*, 1997). EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset.

EML is one of the newer metadatabase standards. The benefit of using EML is that it is a superset of FGDC (eliminating potential incompatibility) and includes a means to describe biological metadata.

To access the EML specification on-line, read it in HTML (Hypertext Markup Language) format, or download the entire specification, including both the HTML documentation and the XML Schema files, browse the EML 2.0.1 Specification at <http://knb.ecoinformatics.org/software/eml/eml-2.0.1/index.html>. The EML modules can be downloaded from the Internet at <http://knb.ecoinformatics.org/software/download.html#eml>; EML XML Schema can be found at <http://www.w3c.org/XML/Schema>, and general information on EML can be found at <http://knb.ecoinformatics.org/software/eml/>.

## 4.3. Dublin Core

The Dublin Core (DC) may be considered to be the best-known metadata project today. From its start in 1995, DC evolved into 'the leading initiative for improving resource discovery on the Web' (Weibel, 2000). During a series of metadata workshops, DC was created and further refined. It is a minimal (15 fields) generic metadata scheme for virtually any kind of document.

DC is a metadata format determined on the basis of international consensus that has defined a minimal information resource description, generally for use in a World-Wide Web environment. The description of the information resources is created using DC elements. An element is a pre-defined string or label, which is paired to a value. In its simplest form it looks like this: 'Author = Ronald Snijder'. A list of fifteen elements is defined, which will be discussed in the next subsection. The syntax of the elements is simple: just the name (or 'Identifier') of the element, and the value of the element. The value may consist of free text or it may be taken from a standardized resource. The description may reside in a separate file or it may be a part of the information resource itself. While no formal syntax rules are defined, several syntax recommendations have been created for generic text files, HTML and the Resource Description Framework (RDF). These are discussed in (Hillmann, 2000).

### 4.3.1. Elements of DC

The DC element set originally consisted of thirteen elements. Below are the elements, as defined in the first DC metadata workshop held in Dublin, Ohio, March 1995. For a detailed discussion of those elements, see Weibel and Godby *et al.* (1995).

- Subject – Topic addressed by the work,
- Title – Name of the object,
- Author – Person(s) primarily responsible for the intellectual content of the object,
- Publisher – Agent or agency responsible for making the object available,
- OtherAgent – Person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work,
- Date – Date of publication,
- ObjectType – Genre of the object, such as novel, poem, or dictionary,
- Form – Data representation of the object, such as Postscript file or Windows executable file,
- Identifier – String or number used to uniquely identify the object,
- Relation – Relationship to other objects,

- Source – Objects, either print or electronic, from which this object is derived, if applicable,
- Language – Language of the intellectual content,
- Coverage – Spatial locations and temporal durations characteristic of the object.

Information on the Dublin Core Metadata Initiative can be found at <http://www.geocities.com/ronaldsnijder/#2>. General information can be found at [http://alexandria.sdc.ucsb.edu/public-documents/metadata/md\\_imp.html](http://alexandria.sdc.ucsb.edu/public-documents/metadata/md_imp.html). The Dublin Core Metadata Glossary can be found at <http://library.csun.edu/mwoodley/dublincoreglossary.html>

#### 4.4. Federal Geographic Data Committee

The Federal Geographic Data Committee (FGDC, <http://www.fgdc.gov/metadata>) Content Standards for Digital Geospatial Metadata (CSDGM) specify the information content of metadata for a set of digital geospatial data. The purpose of the standard is to provide a common set of terminology and definitions for documentation related to these metadata. The standard provides a common set of terminology and definitions for the documentation of geospatial data. The standard establishes the names of data elements and groups of data elements to be used for these purposes, the definitions of these data elements and groups, and information about the values that are to be provided for the data elements. Information about terms that are mandatory, mandatory under certain conditions, and optional (provided at the discretion of the data provider) also is provided by the standard. Description of the metadata elements can be found at the link above.

The FGDC standard has recently been updated to include the Biological Data Profile (<http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf>). The profile broadens the application of the CSDGM so that it is more easily applied to data that are not explicitly geographic (laboratory results, field notes, specimen collections, research reports) but can be associated with a geographic location. The profile changes the conditionality and domain of CSDGM elements, requires the use

of a specified taxonomical vocabulary, and adds elements.

Information on FGDC can be found at National Geospatial Data Clearinghouse. (<http://www.fgdc.gov/metadata>). The FGDC Metadata Glossary can be found at <http://www.fgdc.gov/metadata/csdgm/glossary.html>.

#### 4.5. The ISO 19115 Metadata Standard

The international community, through the International Organization for Standardization (ISO), has developed and approved an international metadata standard, ISO 19115. As a member of ISO, the United States is required to revise the CSDGM in accord with ISO 19115. Each nation can craft their own profile of ISO 19115 with the requirement that it include the 13 core elements. The FGDC is currently leading the development of a U.S. profile of the (ISO) international metadata standard, ISO 19115.

**Table 4.1** ISO 19115 Core Metadata Elements.

| <b>Mandatory Elements</b> | <b>Conditional Elements</b>        |
|---------------------------|------------------------------------|
| Dataset title             | Dataset responsible party          |
| Dataset reference date    | Geographic location by coordinates |
| Dataset language          | Dataset character set              |
| Dataset topic category    | Spatial resolution                 |
| Abstract                  | Distribution format                |
| Metadata point of contact | Spatial representation type        |
| Metadata date stamp       | Reference system                   |
|                           | Lineage statement                  |
|                           | On-line Resource                   |
|                           | Metadata file identifier           |
|                           | Metadata standard name             |
|                           | Metadata standard version          |
|                           | Metadata language                  |
|                           | Metadata character set             |

The status of U.S. Implementation of ISO Metadata can be found at <http://www.fgdc.gov/metadata/geospatial-metadata-standards#isometadata>.

#### 4.6. Metadata stylesheets

Three XML stylesheets have been developed to convert DIF to FGDC and FGDC to DIF. These are listed in Appendix 15.3 and are available from the PICES web site.

#### 4.7. Crosswalks

To reach the broadest community of information workers, metadata must be made available in accordance with a number of popular content metadata standards. As the number, size, and complexity of content metadata standards continues to grow, supplying the metadata for each standard becomes more and more repetitious, time consuming, and tedious. In order to minimize the amount of time needed to create and maintain the metadata and to maximize its usefulness to the widest community of users, there is a need for the metadata created and maintained in one standard to be accessible via related content metadata standards.

A crosswalk is a specification for mapping one metadata standard to another. Technically, a crosswalk is a set of transformations applied to the content of elements in a source metadata standard that results in the storage of appropriately modified content in the analogous elements of a target metadata standard. Crosswalks provide the ability to make the contents of elements defined in one metadata standard available to communities using related metadata standards.

Several crosswalks are available to convert from one metadata standard to another.

- See [http://gcmd.gsfc.nasa.gov/Aboutus/standards/fgdc\\_to\\_dif.html](http://gcmd.gsfc.nasa.gov/Aboutus/standards/fgdc_to_dif.html) to convert from the FGDC to the GCMD DIF standard.
- See [http://gcmd.gsfc.nasa.gov/Aboutus/standards/dublin\\_to\\_dif.html](http://gcmd.gsfc.nasa.gov/Aboutus/standards/dublin_to_dif.html) to convert from DC to the GCMD DIF standard.
- See [http://www.fgdc.gov/metadata/documents/FGDC\\_Sections\\_v40.xls](http://www.fgdc.gov/metadata/documents/FGDC_Sections_v40.xls) to convert from CSDGM to ISO 19115. This tool was developed by Intergraph as part of the NSDI Cooperative Agreement Program (CAP) (<http://www.fgdc.gov/grants/index.html>).
- See [http://gcmd.gsfc.nasa.gov/Aboutus/standards/esri\\_to\\_dif.html](http://gcmd.gsfc.nasa.gov/Aboutus/standards/esri_to_dif.html) to convert from GCMD DIF

format to the Environmental Systems Research Institute, Inc. (ESRI) profile of FGDC.

#### 4.8. Tools for creating metadata

There are many useful tools to assist with creating metadata, some of which are listed below.

- ARC/INFO GIS Metadata Generator AMLs ([ftp://moon.cecer.army.mil/pub/metadata/arc\\_meta/](ftp://moon.cecer.army.mil/pub/metadata/arc_meta/))
- ASCII template ([ftp://moon.cecer.army.mil/pub/metadata/ascii\\_template](ftp://moon.cecer.army.mil/pub/metadata/ascii_template))
- Another ASCII template (<ftp://ftp.nbs.gov/pub/metadata/mptemplt.txt>)
- CORPSMET (<http://sco.wisc.edu/wisclinc/metatool/corpsmet.htm>) is a U.S. Army Corps of Engineers (USACE) program to create metadata.
- MP (<http://geology.usgs.gov/tools/metadata/tools/doc/mp.html>), a compiler for formal metadata by developed by Peter Schweitzer of the United States Geological Survey (USGS).
- NOAA's FGDC Metadata Toolkit Software ([http://www-orca.nos.noaa.gov/cgi-bin/titledetails.pl?2\\_SEA\\_MetadataToolkit](http://www-orca.nos.noaa.gov/cgi-bin/titledetails.pl?2_SEA_MetadataToolkit))
- XTME (<http://geology.usgs.gov/tools/metadata/>) is a Metadata Entry System.
- Metavist 2005 (<http://ncrs.fs.fed.us/pubs/viewpub.asp?key=2737>) is a stand-alone metadata creation and editing tool developed by the USDA Forest Service North Central Research Station. Metavist 2005 creates metadata compliant with the CSDGM (FGDC) 1998 metadata standard and the National Biological Information Infrastructure (NBII) 1999 Biological Data Profile for the FGDC standard. The software runs under the Microsoft Windows 2000 and XP operating systems, and requires the presence of Microsoft's .Net Framework version 1.1. The metadata are output in XML format.