1   Status of this Memo

This memo provides information to the NASA Earth Science Data Systems (ESDS) community.   This memo does not specify an ESDS standard of any kind. Distribution of this memo is unlimited.

2   Change Explanation

Changes made to ESDS-RFC-012v0.1

Added heading/TOC entry for section 7.3; added APA and ALA to Glossary of Acronyms, cleaned up spacing between some paragraphs and sections.

3   Copyright Notice

Copyright © 2008 United States Government as represented by the Administrator of the National Aeronautics and Space Administration. All Rights Reserved.

4   Abstract

This document defines the Directory Interchange Format (DIF) content metadata, a specific set of attributes for describing Earth science data. From its conception in the late 1980's as a way to document and exchange information on scientific data to its implementation in the Global Change Master Directory (GCMD), the DIF has evolved to serve the user community in the discovery, access and use of Earth science and related data.

5   Table of Contents

6   Introduction

The Directory Interchange Format (DIF) is used to create directory entries that describe a group of data. The directory contains several fields, allowing users to find data useful to their needs. The Global Change Master Directory (GCMD) is an implementation of the DIF.

7   Our work

The GCMD database holds more than 20,000 descriptions of Earth science data sets and services covering all aspects of Earth and environmental sciences. One can use the search box or select from the available keywords to search for data and services.

The GCMD offers authoring tools, which conform to international standards. The web-based docBuilder tool allows metadata authors to add or modify data set descriptions (DIFs). Tools are available to write, directly submit, and directly update metadata records. In addition, Subscription services are available to notify users of new entries.

The GCMD offers descriptions of Earth science data sets using a specified set of information - known as the Directory Interchange Format (DIF). The GCMD's philosophy for using the DIF is that the specific set of attributes, the content metadata, is the critical set needed for a user to determine if the data sets returned from a database query are those that define viable data sets for the user's needs. Using a specific set of fields also helps to "normalize" the search for data sets through the use of several alternative search engines.

History of the DIF

Before metadata became a ubiquitous buzzword, a descriptive and standardized format for exchanging information about scientific data sets was conceived and implemented. The Directory Interchange Format, the DIF, was the product of an Earth Science and Applications Data Systems Workshop (ESADS) held February 24-26, 1987 on catalog interoperability (CI). The workshop recommended that a "...first step towards data system interoperability, Catalog Interoperability (CI), the ability to find information about data held at other sites...", be made.

In the summer of 1987, the Catalog Interoperability Working Group (consisting of several U.S. Federal and international agencies) defined the type of information and level of detail that would be contained by the DIF. The DIF structure was "frozen" on September 18, 1987, and the population of NASA's Master Directory (NMD) prototype commenced. By December 1987, over 100 DIF entries were available in the prototype NMD database. After several demonstrations, workshops, and feedback from the scientific community, the Directory Interchange Format (DIF) was formally approved and adopted by a CI science advisory group at a CI workshop in 1988.

In 1989, the Committee on Earth Observation Satellites (CEOS) Data Working Group (DWG) began attending the CI Workshop meetings and provided valuable feedback on the DIF structure. The CEOS International Directory Network (IDN) was soon established under the auspices of the CEOS Working Group on Data, through the Catalog Subgroup, to foster the exchange of information among international agencies. The first release of the IDN was named the Prototype International Directory (PID) in 1990. [Actual DIF exchange procedures were agreed on by February 1991.]

In 1990, the Interagency Working Group on Data Management for Global Change (IWGDMGC) adopted the directory as a prototype to facilitate global change research - in response to the challenge by the Earth System Science Committee (ESSC). Thereafter, the NMD was renamed the Global Change Master Directory (GCMD) for its Earth sciences applications.

The DIF has enjoyed over 16 years of success. The DIF structure has been flexible enough to evolve with growing metadata requirements, especially for the geospatial disciplines. In the U.S., the Office of Management and Budget's (OMB's) Circular A-16 for the improved coordination of spatial data among federal agencies led to the establishment of the Federal Geographic Data Committee (FGDC) and the FGDC Clearinghouse. The GCMD serves as NASA's FGDC Clearinghouse node for geospatial metadata. Elements of the Content Standard for Digital Geospatial Metadata (CSDGM) were incorporated in the DIF in 1994.

In the late 1990s, the geospatial community began work towards the development of an international standard for geospatial metadata. The International Organisation for Standardisation (ISO) Technical Committee (TC 211) Metadata Standard 19115 (previously known as 15046-15) (see: http://www.isotc211.org/) sought to provide "a consistent suite of geographic information schemata that allows geographic information to be integrated with information technology. The goal of this work item is to produce a schema for geographic information metadata."

The ISO 19115/TC 211 geospatial metadata standard was adopted June 2004. Required elements and appropriate modifications were approved by the CEOS IDN Interoperability group and incorporated into the DIF to achieve full ISO compatibility.

The DIF does not compete with other metadata standards. It is simply the "container" for the metadata elements that are maintained in the IDN database, where validation for mandatory fields, keywords, personnel, etc. takes place.

The DIF is used to create directory entries, which describe a group of data. A DIF consists of a collection of fields, with detail specific information about the data. Eight fields are required in the 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 DIF allows users of data to understand the contents of a data set and contains those fields which are necessary for users to decide whether a particular data set would be useful for their needs.

Adding Elements to Satisfy User's Needs

The GCMD willingly adds new elements to meet the needs of different communities. For example, adding the spatial resolution field a few years ago has helped enable users to identify relevant data sets. It was not a request to add this feature but the GCMD staff took it upon themselves to enhance the user's searching experience.

The GCMD hosts and maintains the IPY Metadata Profile input template and repository. Based on DIF, it does not currently contain a field or fields for describing the datum or projection, also called the location coordinate reference system (CRS), for an observation's stated geographic location. The GCMD staff agreed this is an important metadata field and that it should be added to the DIF form for use by the IPY Metadata Profile and any others wishing to use it. The GCMD DIF form has a qualifier for fields on the form, by which a given field could be flagged by the system as "Required if applicable" for data entry consistency checking. The GCMD staff has agreed to incorporate a new CRS field on the metadata form, which would be expected to be populated with the web link or other citation reference to a CRS in a known and recognized CRS registry.

The Interoperability Forum

The Interoperability Forum (Interop Forum), a constantly evolving standard, is designed to share suggestions for the Committee of Earth Observation Satellites' (CEOS) International Directory Network's (IDN) metadata content requirements, syntax specifications, and access

considerations. The goal is to improve the use of the directory through its standards: the Directory Interchange Format (DIF) and the Service Entry Resource Format (SERF). In addition, suggested enhancements, corrections, or other issues may be raised through the Forum, which relate to any aspect of the website, including the use of the docBUILDER authoring tool.
Fields of the DIF

The DIFGuide document, also referred to as the "Writer's Guide," provides information about each field of the DIF, including its syntax, specifications, recommendations, and examples. Several example DIFs are also provided. Information here is meant to be illustrative of the content of the Writer's Guide, which is the sole specification document. The DIF fields included in the Writer's Guide are cited below and are designated as **required**, **highly recommended**, or **recommended**.

These fields are required:

Entry Identifier
The **<Entry_ID>** is the unique identifier of the metadata record. The **<Entry_ID>** is determined by the metadata author and may be identical to identifiers used by the data provider's data center or organization. For example, the National Snow and Ice Data Center (NSIDC) DAAC identifies their metadata records as *NSIDC-xxxx*, where *xxxx* is a numerical designator. Also, the identifier is case insensitive meaning nsidc-xxxx and NSIDC-xxx refer to the same metadata record.

Entry Title
The **<Entry_Title>** is the title of the data set described by the metadata.

<**Entry_Title**> should be descriptive enough so that when a user is presented with a list of titles the general content of the data set can be determined. For example, *<Entry_Title>Aerosols </Entry_Title>* would not be an adequate data set title as it does not provide enough descriptive information to guide the user.

In order to make titles descriptive, important elements about the data may be included in the **<Entry_Title>**, i.e., parameters measured, geographic location, instrument, investigator, project, temporal coverage. For example, *<Entry_Title>Aerosol characterization and snow chemistry at Terra Nova Bay 2001- 2003</Entry_Title>* provides an adequate amount of information to guide the user.

Parameters (Science Keywords)
The **<Parameters>** field allows for the specification of Earth science keywords that are representative of the data set being described. These keywords are important for the precise search and retrieval of information from the GCMD. The author must select these keywords from the controlled set of science keywords. The <**Parameters**> field consists of a 7-level

hierarchical classification of science keywords, defined as follows:

- **<Category>** is the highest keyword category. The default is "*EARTH SCIENCE*".
- **<Topic>** is the next highest level in the keyword hierarchy under **<Category>**.
  There are 14 Topics:
  Agriculture
  Atmosphere
  Biosphere
  Biological Classification
  Climate Indicators
  Cryosphere
  Human Dimensions
  Land Surface
  Oceans
  Paleoclimate
  Solid Earth
  Spectral/Engineering
  Sun-Earth Interactions
  Terrestrial Hydrosphere
- **<Term>** is next level in the hierarchy under **<Topic>**. There can be several controlled **<Term>** keywords for each **<Topic>**. For example, <Topic>Atmosphere</Topic> <Term>Aerosols </Term>.
- **<Variable_Level_1>** is the next level in the hierarchy under **<Term>**. There can be several controlled **<Variable_Level_1>** keywords for each **<Term>**. For example, <Topic>Atmosphere</Topic> <Term>Aerosols </Term> <Variable_Level_1>Aerosol Optical Depth</Variable_Level_1>.
- **<Variable_Level_2>** is the next level in the hierarchy under **<Variable_Level_1>**. There can be several controlled **<Variable_Level_2>** keywords for each **<Variable_Level_1>**. For example, <Topic>Biological Classification</Topic> <Term>Animals/Invertebrates</Term> <Variable_Level_1>Arthropods </Variable_Level_1> <Variable_Level_2>Crustaceans</Variable_Level_2>.
- **<Variable_Level_3>** is the next level in the hierarchy under **<Variable_Level_2>**. There can be several controlled **<Variable_Level_3>** keywords for each **<Variable_Level_2>**. For example, <Topic>Biological Classification</Topic> <Term> Animals/Invertebrates </Term> <Variable_Level_1>Arthropods</Variable_Level_1> <Variable_Level_2>Crustaceans</Variable_Level_2> <Variable_Level_3>Decapods </Variable_Level_3>
- **<Detailed_Variable>** is an uncontrolled free text field that allows the metadata author to specify any keywords to more exactly describe the measurement represented by the data. For example, <Topic>Biological Classification</Topic> <Term> Animals/Invertebrates </Term> <Variable_Level_1>Arthropods</Variable_Level_1>

<Variable_Level_2>Crustaceans</Variable_Level_2>
<Variable_Level_3>Decapods </Variable_Levele_3> <Detailed_Variable>Lobsters
</Detailed_Variable>

ISO Topic Category

The **<ISO_Topic_Category>** field is used to identify the keywords in the ISO 19115 - Geographic Information Metadata (http://www.isotc211.org/) Topic Category Code List. It is a high-level geographic data thematic classification to assist in the grouping and search of available geographic data sets. The **<ISO_TOPIC_Category>** keywords are as follows:

- Farming
- Biota
- Boundaries
- Climatology/Meteorology/Atmosphere
- Economy
- Elevation
- Environment
- Geoscientific Information
- Health
- Imagery/Base Maps/Earth Cover
- Intelligence/Military
- Inland Waters
- Location
- Oceans
- Planning Cadastre
- Society
- Structure
- Transportation
- Utilities/Communications

For definitions of these keywords, see:
http://gcmd.nasa.gov/User/difguide/iso_topics.html

Data Center

The **<Data Center>** is the data center, organization, or institution responsible for distributing the data.
- **<Data_Center_Name>** consists of the data center **<Short_Name>** and **<Long_Name>,** which is the name of the data center that distributes the data.
- **<Data_Center_URL>** is the URL of the data center.
- **<Data_Set_ID>** is a data set identifier assigned by the data center (may or may not be the same as the **<Entry_ID>.**
- **<Personnel>** Contact information for the data including name, email, phone, FAX,

ESDS-RFC-012v1   **Lola Olsen & Chrissy Chiddo**
Category: Standard   **September 2008**
Updates/Obsoletes: None   **Directory Interchange Format (DIF) Standard**

and address information. See the **< Personnel**> entry for a complete description.

Summary

The **<Summary>** is a brief description of the data set that allows potential users to determine if the data set is useful for their needs. The following are suggestions to populate the **<Summary>** field:

- Capitalization should follow standard constructs. For readability, all capital letters or all lower case letters should not be used. Use the appropriate case where applicable.
- Acronyms should be expanded to provide understanding.
- The **<Summary>** may contain tabular information.
- URLs will be automatically hyperlinked.
- Where applicable, the **<Summary>** should include brief statements of the following information:
  - ➢ Data processing information (gridded, binned, swath, raw, algorithms used, necessary ancillary data sets).
  - ➢ Date available.
  - ➢ Data set organization (description of how data are organized within and by file).
  - ➢ Methodology or analytical tools,
  - ➢ Time gaps in data set coverage.
  - ➢ Units and unit resolution.
  - ➢ Similarities and differences of these data to other closely-related data sets.
  - ➢ Purpose and/or intended use of data.
  - ➢ Other pertinent information.

Metadata Name

The ISO 19115 **<Metadata_Name>** field is used to identify the current DIF standard name. This field is automatically populated when using the GCMD metadata authoring tool.

Metadata Version

The **<Metadata_Version>** field is used to identify the current DIF metadata standard. This field is automatically populated when using the GCMD metadata authoring tools.

These fields are highly recommended:

Data Set Citation

The **<Data_Set_Citation>** field allows the author to properly cite the data set producer. This field has 2 functions:

1. To indicate how this data set should be cited in the professional scientific literature, and
2. If this data set is a compilation of other data sets, to document and credit the data sets that

were used in producing this compilation.

This field is not to be used to list bibliographic references of scientific research articles arising from the data set. This field provides a citation for the data set itself, not articles related to the research results. To list references related to the research results, use the **<Reference>** field.

<Data_Set_Citation> consists of:
- **<Dataset_Creator>** The name of the organization(s) or individual(s) with primary intellectual responsibility for the data set's development.
- **<Dataset_Title>** The Title of the data set; this may be the same as Entry Title.
- **<Dataset_Series_Name>**The name of the dataset series or aggregate dataset of which the dataset is a part. For example <Dataset_Series_Name>IGBP PAGES/WDC CONTRIBUTION SERIES</Dataset_Series_Name>
- **<Dataset_Release_Date>** The date when the data set was made available for release.
- **<Dataset_Release_Place>**The name of the city (and state or province and country if needed) where the data set was made available for release.
- **<Dataset_Publisher>**The name of the individual or organization that made the data set available for release.
- **<Version>**The Version of the data set. For example, <Version>Version 1.2</Version>
- **<Issue_Identification>**The volume or issue number of the publication (if applicable).
- **<Data_Presentation_Form>**The mode in which the data are represented, e.g. atlas, image, profile, text, etc.
- **<Other_Citation_Details>**Additional free-text citation information.
- **<Online_Resource>** The URL of the online resource containing the data set.

Personnel
<Personnel> defines the point of contact for more information about the data set or the metadata.
- The contact personnel are defined by the **<Role>,** which include:
  Investigator: The person who headed the investigation or experiment that resulted in the acquisition of the data described (i.e., Principal Investigator, Experiment Team Leader).
  Technical Contact: The person who is knowledgeable about the technical content of the data (quality, processing methods, units, available software for further processing).
  DIF Author: The person who is responsible for the content of the DIF. If the responsibility shifts from the original author to another person, the DIF Author field should be updated to the new responsible person.
- **<First_Name>, <Middle_Name>** and **<Last_Name>** are the first, middle and last name of the person or organization defined in the **<Personnel>** field. Initials may be used for the **<First_Name>** and **<Middle_Name>.** Organizational names may be substituted for personal names.
- **<Email>** is the email address of the personnel or organization. Note: Authors may use "@"

in formatting email addresses, however the "@" symbol will not be displayed to the public to avoid potential spam attacks.
- **<Phone>** is the telephone number of the person or organization. Telephone extensions are allowed. If not in the U.S. or Canada, use the two-digit country code followed by the phone number. Otherwise, use the 10-digit phone number including area code. For example: **<Phone>***+44 5555 555555***</Phone>** or **<Phone>***301-555-5555***</Phone>**
- **<FAX>** is the FAX number of the person or organization. The same rules for **<Phone>** apply to FAX numbers.
- **<Contact Address>** contain the address information of the person or organization. It consists of:
    > **<Address>** is the organization name, department, mail stop, street address, etc. of the person or organization.
    > **<City>** is the city or town of the person or organization.
    > **<Province or State>** is the province (particularly Canadian provinces), region or state (particularly in the United States).
    > **<Postal Code>** is the postal code of the person or organization.
    > **<Country>** is the country of the person or organization.

Instrument (Sensor Name)
The Instrument or **<Sensor_Name>** is the name of the instrument used to acquire the data. There are 3 categories of instruments, with additional levels of hierarchical classifications (see the Writer's Guide for Ancillary Descriptions):
- Earth Remote Sensing Instruments
- In Situ/Laboratory Instruments
- Solar/Space Observing Instruments

When using the GCMD metadata authoring tools, **<Sensor_Name>** classifications are automatically populated by selecting the **<Short_Name>**. This field allows for the specification of keywords that are the names of the instrument(s) used to collect or measure the data. These keywords are important for the search and retrieval of information from the GCMD. The field **<Sensor_Name>** consists of the short name and the long name of the instrument. There is a 1:1 correspondence between the short name and the long name:
- **<Short_Name>** is the abbreviated name of the instrument used to acquire the data.
- **<Long_Name>** is the full name of the instrument used to acquire the data.

Platform (Source Name)
The Platform or **<Source_Name>** is the name of the platform used to acquire the data, There are 11 categories of platforms, with additional levels of hierarchical classifications (see the Writer's Guide for Ancillary Descriptions):
- Aircraft
- Balloons/Rockets
- Earth Observation Satellites

- In Situ Land-based Platforms
- In Situ Ocean-based Platforms
- Interplanetary Spacecraft
- Maps/Charts/Photographs
- Models
- Navigation Platforms
- Solar/Space Observation Platforms
- Space Stations/Manned Spacecraft

When using the GCMD metadata authoring tools, **<Source_Name>** classifications are automatically populated by selecting the **<Short_Name>**. This field allows for the specification of keywords that are representative of the platform(s) or source used to collect the data. These keywords are important for the search and retrieval of information from the GCMD. The field **<Source_Name>** consists of the short name and the long name of the platform (source). There is a 1:1 correspondence between the short name and the long name:
- **<Short_Name>** is the abbreviated name of the platform (source) used to acquire the data.
- **<Long_Name>** is the full name of the platform (source) used to acquire the data.

Temporal Coverage
The **<Temporal_Coverage>** field specifies the start and stop dates during which the data was collected.

Temporal_Coverage> consists of:
- **<Start_Date>** is the starting date of the data collection.
- **<Stop_Date>** is the ending date of the data collection.

Paleo Temporal Coverage
For paleoclimate or geologic data, **<Paleo_Temporal_Coverage>** is the length of time represented by the data collected. **<Paleo_Temporal_Coverage>** should be used when the data spans time frames earlier than yyyy-mm-dd = 0001-01-01.

<Paleo_Start_Date> is the number of years furthest back in time including units Ga, Ma, ka or ybp.
<Paleo_Stop_Date> is the number of years closest to the present time including units Ga, Ma, ka or ybp.
<Chronostratigraphic_Unit> is a controlled keyword describing Geologic Time of the form:
- **<Eon>** is the primary and largest division of geologic time. Limits correspond with major, global crustal events, changes in sea level and/or climate, or biotic changes.
- **<Era>** is a major division of geological time, tens or hundreds of millions of years long, usually distinguished by significant changes in the plant and animal kingdoms; also used to denote later archaeological periods, such as the prehistoric era. Five geologic eras have

been established: Archeozoic [before 2500 million years ago (Ma)], Proterozoic (2500 to 570 Ma), Paleozoic (570 Ma to ca. 250 Ma), Mesozoic (ca. 250 to ca. 70 Ma), and Cenozoic (since ca. 70 Ma). All eras are divided into at least two geologic periods.

- **<Period>** is the secondary division of geologic time, delimited by moderate but usually global crustal events, changes in sea level and/or climate, or biotic changes, sometimes in a relatively localized area. Two or more periods are required to make up a geologic era, and each period comprises two or more geologic epochs.
- **<Epoch>** is the third-order division of geologic time, delimited by partial withdrawal of the sea from land masses and by gentle crustal disturbances in localized areas. Two or more epochs are required to make up a geologic period, and, in turn, two or more periods are needed to constitute a geologic era.
- **<Stage>** is a smaller unit of time. An Epoch usually has 1 or more Stages.

Spatial Coverage

The **<Spatial_Coverage>** field specifies the geographic and vertical (altitude, depth) coverage of the data.

- **<Southermost_Latitude>** The southernmost geographic latitude covered by the data. From: 0 - 90° North or 0 - 90° South.
- **<Northernmost_Latitude>** The northernmost geographic latitude covered by the data. From: 0 - 90° North or 0 - 90° South.
- **<Westernmost_Longitude>** The westernmost geographic longitude covered by the data. From: 0 - 180° east or 0 - 180° west. The Prime Meridian is 0 degrees, measured positive (+) eastwards of the PM.
- **<Easternmost_Longitude>** The easternmost geographic longitude covered by the data. From: 0 - 180° east or 0 - 180° west. The Prime Meridian is 0 degrees, measured positive (+) eastwards of the PM.
- **<Minimum_Altitude>** The altitude level, which represents the lower limit of data coverage. Units can be specified.
- **<Maximum_Altitude>** The altitude level, which represents the higher limit of data coverage. Units can be specified.
- **<Minimum_Depth>** The depth level, which represents the upper-most depth of data coverage. Units can be specified.
- **<Maximum_Depth>**The depth level, which represents the lowest depth of data coverage. Units can be specified.

Location

The **<Location>** field specifies the name of a place on Earth, a location within the Earth, a vertical location, or a location outside of Earth. The **<Location>** keywords are a 5-level hierarchy of controlled keywords under the following **Categories**:

- **Continent**: specifies continents:

➢ Africa
➢ Antarctica
➢ Asia
➢ Australia/New Zealand
➢ Europe
➢ North America
➢ South America

• **Ocean**: specifies an ocean body:
➢ Arctic Ocean
➢ Atlantic Ocean
➢ Indian Ocean
➢ Pacific Ocean
➢ Southern Ocean

• **Geographic Region**: specifies a region on Earth (e.g., Polar):
➢ Arctic
➢ Eastern Hemisphere
➢ Equatorial
➢ Eurasia
➢ Global
➢ Global Land
➢ Global Ocean
➢ Mid-latitude
➢ Northern Hemisphere
➢ Oceania
➢ Polar
➢ Sahel
➢ Southern Hemisphere
➢ Tropics
➢ Western Hemisphere

• **Solid Earth**: specifies a region within the Earth:
➢ Core
➢ Crust
➢ Mantle

• **Space**: specifies a region outside of Earth, particularly regions of Earth-Sun interactions:
➢ Earth Magnetic Field
➢ Solar Region

- **Vertical Location**: specifies a region in the vertical direction, either in the atmosphere or under the oceans:
  - ➢ Boundary Layer
  - ➢ Ionosphere
  - ➢ Land Surface
  - ➢ Mesosphere
  - ➢ Sea Floor
  - ➢ Sea Surface
  - ➢ Stratosphere
  - ➢ Thermosphere
  - ➢ Troposphere

Data Resolution

The **<Data_Resolution>** field specifies the resolution of the data, which is the difference between two adjacent geographic, vertical, or temporal values. Controlled keywords representing horizontal, vertical and temporal data resolution ranges can be selected. Selection of data resolution ranges will assist users in refining their search for data within specific resolution ranges.

- **<Latitude_Resolution>** is the minimum difference between two adjacent latitude values. For example, <Latitude_Resolution>2.5 degrees </Latitude_Resolution>
- **<Longitude_Resolution>** is the minimum difference between two adjacent longitude values.
- **<Horizontal_Resolution_Range>** is a controlled list of latitude/longitude resolution ranges. The range should be selected based on the **<Latitude_Resolution>** and **<Longitude_Resolution>**.
- **<Vertical_Resolution>** is the minimum difference possible between two adjacent vertical values.
- **<Vertical_Resolution_Range>** is a controlled list of vertical resolution ranges. The range should be selected based on the specified **<Vertical_Resolution>**.
- **<Temporal_Resolution>** is the frequency of data sampled.
- **<Temporal_Resolution_Range>** is a controlled list of temporal resolution ranges. The range should be selected based on the specified **<Temporal_Resolution>**.

Project

The **<Project>** is the name of the scientific program, field campaign, or project from which the data were collected.

This field allows for the specification of keywords that are the names of the projects. These keywords are important for the search and retrieval of information from the GCMD. The field **<Project>** consists of the short name and the long name of the project. There is a 1:1 correspondence between the short name and the long name:

- **<Short_Name>** is the abbreviated name of the project from which the data were collected.

ESDS-RFC-012v1
Category: Standard
Updates/Obsoletes: None

Lola Olsen & Chrissy Chiddo
September 2008
Directory Interchange Format (DIF) Standard

- **<Long_Name>** is the full name of the project from which the data were collected.

Quality

The **<Quality>** field allows the author to provide information about the quality of the data
or any quality assurance procedures followed in producing the data described in the metadata.
Suggestions for information to include in the **<Quality>** field:
- Description should be succinct.
- Include indicators of data quality or quality flags.
- Include recognized or potential problems with quality.
- Established quality control mechanisms should be included.
- Established quantitative quality measurements should be included.

Access Constraints

The **<Access_Constraints>** field allows the author to provide information about any constraints
for accessing the data set. This includes any special restrictions, legal prerequisites, limitations
and/or warnings on obtaining the data set. Some words that may be used in this field include:
Public, In-house, Limited. Additional detailed instructions on how to access the data can be
entered in this field.

Use Constraints

The **<Use_Constraints>** field allows the author to describe how the data may or may not be
used after access is granted to assure the protection of privacy or intellectual property. This
includes any special restrictions, legal prerequisites, terms and conditions, and/or limitations on
using the data set. Data providers may request acknowledgement of the data from users and
claim no responsibility for quality and completeness of data.

Distribution

The **<Distribution>** field describes media options, size, data format, and fees involved in
distributing the data set.
- **<Distribution_Media>**The media options for the user receiving the data.
- **<Distribution_Size>** An approximate size (in KB, MB or GB) for the entire data set.
  Specify if data are compressed and the method of compression.
- **<Distribution_Format>** The data format used to distribute the data.
- **<Fees>**Cost of **<Distribution_Media>** or distribution costs if any. Specify if there are no
  costs.

Data Set Language

<Data_Set_Language> describes the language used in the preparation, storage, and description
of the data. It is the language of the information object, not the language used to describe or
interact with the metadata record. **<Data_Set_Language>** does not refer to the language of the
metadata.

Data Set Progress

The **<Data_Set_Progress>** describes the production status of the data set regarding its completeness. There are three choices:
- **Planned** refers to data sets to be collected in the future and are thus unavailable at the present time. For Example: The Hydro spacecraft has not been launched, but information on planned data sets may be available.
- **In Work** refers to data sets currently undergoing production or data that is continuously being collected or updated. For Example: data from the AIRS instrument on Aqua is being collected continuously.
- **Complete** refers to data sets in which no updates or further data collection will be made. For Example: Nimbus-7 SMMR data collection has been completed.

Related URL

The **<Related_URL>** field specifies links to Internet sites that contain information related
to the data, as well as related Internet sites such as project home pages, related data archives/servers, metadata extensions, online software packages, web mapping services, and calibration/validation data. The **<Related_URL>** field consists of:
- **<URL_Content_Type>** describes the type of URL being referenced. The **<URL_Content_Type>** is selected from a list of controlled URL content type keywords and consists of:
  <Type> describes the type resources being referenced by the URL.
  <Subtype> describes the subtype of the resource being referenced by the URL.
- **<URL>** is the URL to the resource associated with the data set.
- **<Description>** provides information about the resource defined by the **<URL>**.

DIF Revision History

The **<DIF_Revision_History>** allows the author to provide a list of changes made to the DIF over time. This provides a mechanism for tracking revisions to DIF content.

These fields are optional, but recommended:

Keyword (Ancillary Keyword)

The **<Keyword>** field allows authors to provide any words or phrases needed to further describe the data set.

Originating Center

The **<Originating_Center>** is the data center or data producer who originally generated the dataset.

Multimedia Sample

The **<Multimedia_Sample>** field allows the author to provide information that will enable the display of a sample image, movie or sound clip within the DIF.

- **<File>** describes the filename where the multimedia sample can be found. If the ultimedia sample file resides on the GCMD server, the filename should be specified in this field. If the file is to be transferred to the GCMD, arrangements should be made with GCMD staff.(Note: this is not typical).
- **<URL>** is the URL to be accessed.
- **<Format>** is the format of the multimedia sample file (which may differ from the format of the data files), i.e. GIF, TIFF, JPEG.
- **<Caption>** is a one-line description of the multimedia sample used as a caption when the sample is displayed. The caption is especially useful for images such as graphs and photos.
- **<Description>** A more detailed description of the multimedia sample.

Reference (Publications/References)
The <**Reference**> field describes key bibliographic citations pertaining to the data set. Bibliographic citations may be provided in styles used by professional scientific journals such as APA or MLA.

Parent DIF
The **<Parent_DIF>** field allows the capability to relate generalized aggregated metadata records (parents) to metadata records with highly specific information (children). Population of the **<Parent_DIF>** field should be reserved for instances where many metadata records are basically subsets that can be better represented by one parent metadata record, which describes the entire collection. Typically, the parent metadata record will have many children metadata records, which refer to the parent through the **<Parent_DIF>** field. In some instances, a child may point to more than one parent. The **<Parent_DIF>** is populated with an **<Entry_ID>.**

The **<Parent_DIF>** is only populated if there are children metadata records that reference a parent metadata record. Upon display of results, both parent and children metadata are available with explicit linkages between parent and child records through the **<Parent_DIF>** field.

IDN Node
The Internal Directory Name (IDN) Node **(<IDN_Node>)** field is used internally to identify association, responsibility and/or ownership of the dataset, service or supplemental information. Note: The **<IDN_Node>** field is usually not displayed to the user. The author may populate <**IDN_Node>** from a set of controlled keywords.

DIF Creation Date
The **<DIF_Creation_Date>** specifies the date the metadata record was created. The **<DIF_Creation_Date>** is automatically populated by the authoring tools with the date the metadata entry was created; however, the author may change the date.

Last DIF Revision Date

The **<Last_DIF_Revision_Date>** specifies the date the metadata record was created. The **<Last_DIF_Revision_Date>** is automatically populated by the authoring tools with the date the metadata entry was created or modified; however, the author may change the date.

Future DIF Review Date

The **<Future_DIF_Revision_Date>** allows for the specification of a future date at which the DIF should be reviewed for accuracy of scientific or technical content. Future metadata changes may be due to:
 • Anticipated changes in the personnel fields.
 • Planned changes to the data content (e.g., change in processing algorithm).
 • Expected inclusion of the data as part of a campaign.
 • Planned review of metadata and/or data by the data producer.

Privacy Status

The **<Private>** field allows the author to restrict the data set description from being publicly available. There are only two options:
 • **True**. Make the description unavailable to the public.
 • **False**. Make the description publicly available. This is the default. If the author wants the description to be available to the public, then no action is required.

7.1   Additional Work

Science Keywords

The GCMD permits the normalization of the search for DIFs through 14 sets of controlled keywords, counting Platforms and Instruments. Earth science keywords evolve as new metadata records are added and existing records are modified to meet the changing needs of the Earth science community.

The following rules have been used in determining TERMs (i.e. **<Term>** level keywords**)** and three levels of VARIABLEs (i.e. **<Variable_Level_1>**, **<Variable_Level_2>**, **<Variable_Level_3>** level keywords) for the GCMD keywords. These rules are used in GCMD's procedures for modifying TOPICs (i.e. **<Topic>** level keywords), TERMs, and VARIABLEs to assist the user in locating Earth science data sets of interest. These VALIDs are expected to remain fairly stable over time, although suggestions for additions and/or changes will always be considered. In addition, an uncontrolled level of keywords is available for "detailed variables". This list will be uncontrolled except for spelling. Data set producers and DIF writers are encouraged to populate this field with "detailed variables" for the data sets being documented. None of the rules below will apply to this uncontrolled set of "detailed variables". [The field will be searchable through a Lucene fielded search.] Be aware that not all the keywords currently have dataset descriptions behind them.

18

1. At any level of the keyword taxonomy, all TOPICs, TERMs, and VARIABLEs should be chosen to be mutually exclusive, minimizing overlap as much as possible.

2. At any level within the taxonomy, the keywords should be parallel. For example, one would not include a broader or narrower keyword within any one level of the taxonomy.

3. Terms may be prefixed with TOPIC level modifiers, if they do not "stand alone" well.

4. TERMs/VARIABLEs should be plural when singular vs. plural is in question.

    Count nouns answering the question, "How many?" are plural: for example, chemical reactions, penguins, ecosystems. (Exceptions to this rule exist on a discipline-specific basis.)

    Non-count nouns answer the question "How much?" Abstract concepts and unique entities are singular: for example, copper, snow, water, digestion, and conductivity.

5. No "data center-specific" pre- or suf-fixed VARIABLEs should be used.

6. Chemical symbols may be used at **<Variable_Level_2>** or **<Variable_Level_3>** or as detailed variables.

7. VARIABLEs may be prefixed with TERM level modifiers; however, this is not required. TERM modifiers are suggested to identify VARIABLEs that do not "stand alone" well. A generic VARIABLE such as "motion" should not be used if a more accurate and descriptive VARIABLE such as "sea ice motion" is what the user will find in the search. [VARIABLEs should generally not be prefixed with TOPIC level modifiers, although there are exceptions]

8. Statistical modifiers may be used at **<Variable_Level_2>** or **<Variable_Level_3>** or captured as detailed VARIABLEs.

    Example:
    Mean stream discharge

9. Extended modifiers should be reserved for **<Variable_Level_2>** or **<Variable_Level_3>** or captured in the uncontrolled (detailed VARIABLE) keyword list.
    Example:
    Integrated Precipitable Water Vapor
    "Intercepted" Photosynthetically Active Radiation

10. Meaningless (scientifically,) overly complex modifiers, or internal organization prefixes should be avoided in the VARIABLE list.

    Example:
    "1_BUTENE" and "Langley_8_year_SRB_SW_Radiation" would be appropriate only at the uncontrolled detailed VARIABLE list.

11. If a generic VARIABLE has been used to describe a contingent of VARIABLEs, a repetitive generic TERM should NOT be used at the same level of the VARIABLE list. When multiple expressions for the same VARIABLE exist, the VARIABLE level should indicate these as "xxx Expressions", signifying that the expressions following can be used interchangeably with appropriate conversions.

    Example:
    Variable = water vapor. Do not include another VARIABLE indicating the same quantity at the same level, such as "humidity". All water vapor derived values (which can be converted from one expression to another), such as "absolute humidity", "specific humidity", "relative humidity", "vapor pressure", "mixing ratio" should be listed at the VARIABLE level below the common identifier, "Water Vapor Expressions".

12. Duplicate VARIABLEs should be avoided if one serves as a euphemism or surrogate for another.

    Possible example:
    sea ice stage development vs. sea ice form

13. Variable descriptors that add only nebulous information should be avoided; for instance, how low is low? for the lower troposphere?

14. If the science community uses terms interchangeably, the more commonly used variable in the field should be chosen.

    Examples:
    VARIABLE: Use Planetary Boundary Layer (PBL) or Atmospheric Boundary Layer (ABL)
    Reserve "peplosphere" for the Detailed VARIABLE level

15. Modifiers that only describe the spatial domain should generally be reserved for **<Variable_Level_2>** or **<Variable_Level_3>**.

    Example:
    <Variable_Level_1>: Heat Flux

<Variable_Level_2> or **<Variable_Level_3>**: Global Heat Flux

16. VARIABLEs should be mutually exclusive, minimizing overlap as much as possible.

17. Keywords should not be associated with "value judgments". "Air Quality" should be used in preference to "Pollution".

18. Any "slashed" keywords must be clarified so that each side of the slashed word(s) can stand-alone for searching by the user.

    Example:
    Use: Atmosphere > Atmospheric Radiation > Optical Depth/Optical Thickness
    Not: Atmosphere > Atmospheric Radiation > Optical Depth/Thickness

19. An overriding goal is to have the keywords as "parallel" (in terms of "detail") as possible within any one level of the hierarchy.

Implementations

The DIF is being used by the Joint Committee on Antarctic Data Management (JCADM), an international effort to make available Antarctic research data holdings. The GCMD has developed an Antarctic Master Directory (AMD) to search for Antarctic data sets.
The GCMD is also an NSDI/FGDC participant and a node of the FGDC Clearinghouse. Geospatial data can be searched across many different spatial servers by accessing the FGDC Clearinghouse Gateway.

7.2   Collaborations (i.e. Who's using the DIF)

NASA Collaborations

•   **NASA Earth Observing System Data Information System (EOSDIS)**
    The GCMD provides data set descriptions for all of NASA's EOSDIS publically available data holdings. Users can link directly to the EOSDIS Data Gateway (EDG) or to data held at the Distributed Data Archive Centers (DAACs) from the data set or services descriptions.

•   **Federation of Earth Science Information Partners (ESIP)**
    The GCMD collaborates directly with two Federation of Earth Science Information Partners (ESIP) and indirectly with others. ESIP partners include OPeNDAP: Open-source Project for a Network Data Access Protocol and the Seasonal and Interannual

Information Partnership (SIESIP). GCMD has worked closely with DODS/OPeNDAP to provide direct access to DODS servers including the Live Access Server (LAS) and to provide DODS direct access to GCMD's database through the GCMD Open API. A custom DODS/OPeNDAP portal is available. The GCMD has provided a customized ESIP data portal and ESIP services portal to search for data. GCMD is also searchable from the ESIP Federation Interactive Network for Discovery (FIND).

U.S. Government Federal Agency Collaborations

- **Biological Resources Division, US Geological Survey**
  The Biological Resources Division (BRD) provides the scientific understanding and technologies needed to support the sound management and conservation of U.S. biological resources. NASA and the BRD are partnering to increase access to ecological data through shared metadata population efforts of the GCMD and the National Biological Information Infrastructure (NBII).

- **Geospatial One-Stop (GOS)**
  Metadata records about NASA's Earth science data are being contributed from the GCMD to the Geospatial One-Stop (GOS) portal. GOS is a U.S. Office of Management and Budget (OMB) E-Government initiative that allows for improved access to geospatial information. Geospatial One Stop's harvesting technology allows for continuous updates of metadata holdings in their database.

- **National Oceanic and Atmospheric Administration (NOAA)**
  GCMD continues close collaborations with NOAA through the exchange of metadata information. The GCMD is a participant in the NOAA Operational Model Archive and Distribution System (NOMADS) program and has contributed a custom portal for model output data sets. Over 800 metadata records have been exchanged with the NOAA National Geophysical Data Center (NGDC) and the National Climatic Data Center (NCDC) using the OAI-PMH standard.

Intergovernmental Collaborations

- **National Spatial Data Infrastructure/Federal Geographic Data Committee (NSDI/FGDC)**
  NSDI is an umbrella of policies, standards and procedures (coordinated by the Federal Geographic Data Committee) under which organizations and technologies interact to foster more efficient use, management and production of geospatial data. The GCMD is an NSDI/FGDC participant and a node of the FGDC Clearinghouse. Geospatial data can

be searched across many different spatial servers by accessing the FGDC Clearinghouse Gateway.

- **Data Management and Communications (DMAC)**
  The GCMD has contributed to the U.S. Integrated Ocean Observing Program (IOOS) proposed Data Management and Communications (DMAC) system. The GCMD will participate in the search and discovery of IOOS data sets.

- **Marine Metadata Interoperability (MMI)**
  The GCMD is collaborating with the Marine Metadata Interoperability (MMI) project, a community effort aimed at making marine science data sets easier to find, access, and use. Scientists and data managers can find valuable information about data management and metadata policies on the MMI website. The GCMD contributes by sharing its extensive list of keywords and metadata expertise with the community.

International Collaborations

- **Committee on Earth Observation Satellites International Directory Network (CEOS IDN)**
  The CEOS IDN is an international effort to assist researchers in locating information on available data sets. The GCMD shares information with other CEOS IDN nodes throughout the world and also provides software, search interfaces and metadata writing tools.

- **Antarctic Master Directory**
  The GCMD is collaborating with the Joint Committee on Antarctic Data Management (JCADM), an international effort to make available Antarctic research data holdings. The GCMD has developed an Antarctic Master Directory (AMD) to search for Antarctic data sets.

- **United Nations Environmental Programme (UNEP)**
  The GCMD is continuing collaborations with the United Nations Environmental Programme (UNEP) Global Resources Information Database (GRID) to make available through data set descriptions from many nations throughout the world. The UNEP GRID site in Budapest, Hungary is a partner in the testing of new GCMD software technologies.

- **The Global Observing Systems Information Center (GOSIC)**
  The GCMD is collaborating with the Global Observing Systems Information Center (GOSIC) to develop a portal to GOSIC data that will be collected by three major observing programs: the Global Climate Observing System (GCOS), the Global Ocean

Observing System (GOOS), and the Global Terrestrial Observing System (GTOS). The GCMD has contributed a customized Global Observation of Forest Cover (GOFC) portal GOFC, originally a joint program of the Committee on Earth Observation Satellites (CEOS) and GTOS, now is part of the GTOS program.

- **The International Oceanographic Data and Information Exchange (IODE)**
  GCMD has collaborated with the International Oceanographic Data and Information Exchange (IODE) by testing the MEDI metadata tool. MEDI is compatible with the Directory Interchange Format (DIF) and IODE has encouraged their data centers to create new MEDI records and send metadata to the GCMD.

- **Global Ocean Ecosystems Dynamics Project (GLOBEC)**
  The GCMD has continued collaborations with the Global Ocean Ecosystems Dynamics Project (GLOBEC) to make GLOBEC data sets available through the GCMD and a GLOBEC portal.

- **The Gulf Of Maine Ocean Data Partnership (GoMODP)**
  The GCMD is collaborating with the Gulf of Maine Ocean Data Partnership (GoMODP) to promote and coordinate the sharing, linking, and use of data in the Gulf of Maine region. Data sets from the region are made available through the GoMODP portal.

- **The Ocean Biogeographic Information System (OBIS)**
  The Ocean Biogeographic Information System (OBIS) is the information component of the Census of Marine Life (CoML). It is a web-based provider of global geo-referenced information on marine species. Marine species data collections from all of the world's oceans can be discovered through the OBIS portal.

- **International Polar Year (IPY)**
  With the beginning of the International Polar Year (2007/2008) in March 2007, the GCMD is participating in a worldwide effort in the data and information management of IPY.   The directory's Expression of Intent (EoI) for an IPY Metadata and Information Portal Network for the Data and Information Service (DIS) is in development. The GCMD EoI along with other related EoIs were brought together under the IPY DIS for Distributed Data Management full proposal submitted to NSF in April 2006. The JCADM/AMD and ARCTIC portal contributions are expected to play a key role in IPY. In April 2006, a prototype IPY portal was established.

- **Integrated Ocean Observing System (IOOS)**
  The GCMD is collaborating with the Integrated Ocean Observing System (IOOS) as a member of the expert IOOS metadata team. GCMD actively participates in bimonthly telecons to assist with identifying potential metadata requirements (e.g. Content

requirements, vocabularies and ontologies, and discovery mechanisms) for use in the IOOS Data Management and Communications (DMAC) system.

- **Global Earth Observation System of Systems (GEOSS)**
  The Global Earth Observation System of Systems (GEOSS) is an international effort to monitor the Earth for societal benefits and sustainable development. GCMD has developed web portals that will potentially increase the use of Earth observation data and information. GCMD continues to coordinate with the global community of international (e.g. CEOS International Directory Network), intergovernmental, and regional organizations providing solutions for the 10-year GEOSS Implementation Plan.

- **International Heliosphysical Year (IHY)**
  The Primary Objectives of International Heliosphysical Year (IHY) include: Advancing the Understanding of the Heliophysical Processes that Govern the Sun, Earth and Heliosphere Continuing the tradition of international research and advancing the legacy of the International Geophysical Year Demonstrating the Beauty, Relevance and Significance of Space and Earth Science to the World.

- **Electronic Geophysical Year (eGY)**
  The Electronic Geophysical Year (eGY) provides an opportunity for the international geoscientific community to focus effort on a 21st Century e-Science approach to issues of data stewardship: open access to data, data preservation, data discovery, data rescue, capacity building, and outreach. The development of Virtual Observatories and Laboratories is a central feature of eGY.

  eGY is an internationally-recognized resolve by the science community to achieve a step increase in making past, present, and future geoscientific data readily, rapidly, conveniently, and openly available. eGY provides the international framework and a target for stimulating and coordinating activities to make this happen. eGY focuses on themes of electronic data location and access, permission and release of data, conversion of data into modern digital form, data preservation, capacity building, particularly in developing countries, and outreach. Promoting the development of a network of virtual observatories is a central feature of eGY.

Portal Collaborations

GCMD has recognized the importance of customization for partner organizations and is generating subset views of the GCMD directory through portals. Portals have made it easier for

organizations to maintain and document their data in one place without duplicating the effort to create another online directory.

Many organizations acknowledge the importance of metadata related to their area of interest but do not have the resources required to manage the content. One possible solution to this dilemma is to host the metadata through the GCMD and create a portal to view the virtual subset of the metadata (DIFs). Portals help provide science, or application-specific focus for other agencies, science focus groups, consortia, etc. Portals may be trademarked with the logo of an organization while possessing the full functionality of the GCMD search engine and tools.

Using a portal to search a virtual subset is advantageous in that as metadata (DIFs) are added to the subset, it is also freely available from the GCMD general search pages for scientists in other disciplines to access and use.

Usage statistics are also regularly tracked and available. Over one hundred portals are already in use. Please see GCMD Portals. Some portals in the GCMD are not publicly accessible, only for internal use by specific organizations or user groups, upon request. Currently free-text and/or a keyword interface are offered.

7.3   GCMD Homepage & DIF DocBuilder Tool

ESDS-RFC-012v1
Category: Standard
Updates/Obsoletes: None

Lola Olsen & Chrissy Chiddo
September 2008
Directory Interchange Format (DIF) Standard

Figure 1. GCMD Homepage

Figure 2. DIF DocBUILDER

8   References

Normative References

1. Directory Interchange Format (DIF) Writer's Guide, 2008. Global Change Master Directory. National Aeronautics and Space Administration. [http://gcmd.nasa.gov/User/difguide/].

2. Olsen, L.M., G. Major, K. Shein, J. Scialdone, R. Vogel, S. Leicester, H. Weir, S. Ritz, T. Stevens, M. Meaux, C.Solomon, R. Bilodeau, M. Holland, T. Northcutt, R. A. Restrepo, 2007. NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 6.0.0.0.0

Informative References

1. GCMD Main Web Page: http://gcmd.nasa.gov/index.html

2. GCMD Data Sets docBuilder:
    http://gcmd.nasa.gov/DocumentBuilder/Home.do?Portal=GCMD&MetadataType=0

3. GCMD Portal Index: http://gcmd.nasa.gov/Data/portal_index.html

9   Authors' Address

Lola M. Olsen, GCMD Project Manager
Goddard Space Flight Center, Greenbelt, MD 20770
Phone: 301-614-5361, Fax: 301-614-5268, email: Lola.M.Olsen@nasa.gov

10 Appendix A

Glossary of acronyms

Acronym	Description
AMD	Antarctic Master Directory
APA	American Psychological Association
BRD	Biological Resources Division, USGS
CI	Catalog Interoperability
CCSDS	Consultative Committee for Space Data Systems
CEOS	Committee on Earth Observation Satellites
CoML	Census of Marine Life
CRS	Coordinate Reference System
CSDGM	Content Standard for Digital Geospatial Metadata
DAACs	Distributed Data Archive Centers, NASA
DIF	Directory Interchange Format
DIS	Data and Information Service, IPY
DMAC	Data Management and Communications system, IOOS

DODS	Distributed Ocean Data System
DWG	Data Working Group, CEOS
ECS	EOSDIS Core System
EDW	EOSDIS Data Gateway
eGY	Electronic Geophysical Year
EoI	Expression of Intent
EOSDIS	Earth Observing System Data and Information System
ESADS	Earth Science and Applications Data Systems Workshop
ESIP	Federation of Earth Science Information Partners
FGDC	Federal Geographic Data Committee
FIND	Federation Interactive Network for Discovery, ESIP
GCMD	Global Change Master Directory
GCOS	Global Climate Observing System
GEOSS	Global Earth Observation System of Systems
GLOBEC	Global Ocean Ecosystems Dynamics Project
GOFC	Global Observation of Forest Cover
GoMODP	Gulf of Maine Ocean Data Partnership
GOOS	Global Ocean Observing System
GOS	Geospatial One-Stop
GOSIC	Global Observing Systems Information Center
GRID	Global Resources Information Database
GTOS	Global Terrestrial Observing System
IDN	International Directory Network, CEOS
IHY	International Heliosphysical Year
IODE	International Oceanographic Data and Information Exchange
IOOS	Integrated Ocean Observing System
IPY	International Polar Year
ISO	International Organisation for Standardisation
IWGDMGC	Interagency Working Group on Data Management for Global Change
JCADM	Joint Committee on Antarctic Data Management

MLA	Modern Language Association of America
MMI	Marine Metadata Interoperability
NASA	National Aeronautics and Space Administration
NBII	National Biological Information Infrastructure
NMD	NASA's Master Directory
NOAA	National Oceanic and Atmospheric Administration
NOMADS	NOAA Operational Model Archive and Distribution System
NSDI	National Spatial Data Infrastructure
NSF	National Science Foundation
NSIDC	National Snow and Ice Data Center
OAI-PMH	Open Archives Initiative-Protocol for Metadata Harvesting
OBIS	Ocean Biogeographic Information System
OMB	Office of Management and Budget
SIESIP	Seasonal and Interannual Information Partnership
TC 211	International Organisation for Standardisation (ISO) Technical Committee
UNEP	United Nations Environmental Programme
USGS	U. S. Geological Survey