Brief Guide to DIF-10 (Directory Interchange Format-Version 10)

[updated to reflect changes made in DIF-10.2]

Introduction:

The DIF-9 format has been updated to comply with the new Unified Metadata Model for Collection Metadata (UMM-C) fields for the Common Metadata Repository (CMR)[1]. The current format used for metadata exchange with the GCMD is DIF-9 (Directory Interchange Format-Version 9). The DIF-9 format will continue to be supported.

The summary table below shows the major changes included in the DIF-10 format. (Any new field or change to a required, highly recommended or recommended field is considered to be a major change.) The changes described below were made to: (1) conform with the UMM-C model[2], and (2) add fields from ECHO10[3] to limit the loss of information when merging the ECHO and DIF formats.

Three types of changes are described in the table below: 1) new fields, 2) fields whose status (required, highly recommended, recommended) has changed, and 3) fields that have been re-named, re-structured or removed.

The complete DIF-10 schema is available at:

https://git.earthdata.nasa.gov/projects/EMFD/repos/dif-schemas/browse/10.x/dif_v10.2.xsd

---

[1] The CMR will provide access to all ESDIS metadata holdings in a single repository. CMR documentation is available on the Earthdata Wiki at the following URL: https://wiki.earthdata.nasa.gov/display/CMR/Common+Metadata+Repository+Home

[2] The UMM-C model (version 2) is currently under review by NASA's ESDIS Standards Office (ESO). UMM-C schema documentation is available on the Earthdata Wiki at the following URL: https://wiki.earthdata.nasa.gov/display/CMR/Internal+Team+Documentation

[3] ECHO10 collection level schema is available at the following URL: https://api.echo.nasa.gov/ingest/schema/Collection.xsd

Changes in DIF-10:

DIF-10 Field	Field Definition	Change in DIF-10
New Fields		
Version_Description	A brief description of the differences between one data set version and another version.	The Version Description field allows for a description of what makes one instance of a product different from other versions.
Metadata_Association	Describes the metadata associated with the data set; i.e., the name and other details of input data, data sets associated (in science data terms) with the instance and/or data sets dependent on the collection.	The Metadata Association field provides for parent-child linkages, data sets used in the generation of the collection, or dependent collections of the data set. The "Parent Metadata" field is now a sub-field under Metadata Association.
Metadata_Dates	A union of the DIF metadata event date fields with the three ECHO event time fields.	The Metadata Dates field allows authors to choose either a date or datetime field depending on the known or desired data resolution.
Additional_Attributes	Parameters which further describe the data represented in each granule within a collection.	The Additional Attributes field allows users to provide product specific attributes such as: data type, measurement resolution, parameter range, parameter unit of measure and parameter value accuracy.
Product_Level_Id	The product identifier of the data collection.	The Product Level Id field allows for specifying product specific information, such as data set name, version, product level, etc.
Collection_Data_Type	Identifies non-science-quality products.	The Collection Data Type field specifically identifies the type of data as: "Near Real Time" or "Science" quality.
Fields with Change in Status		
Project	The name of the scientific program, field campaign, or project from which the data were collected.	The Project field is now a required field. In addition, Campaign is a new sub-field under Project and start and end dates can also be provided.

Temporal_Coverage	Specifies the start and stop dates during which the data were collected.	Temporal Coverage is now a required field. This field has also been updated to offer additional methods to define temporal information. In addition, Paleo-Temporal Coverage has been moved within this group to provide one location for all temporal information.
Spatial_Coverage	Specifies the geographic and vertical (altitude, depth) coverage of the data.	Spatial Coverage is now a required field. It has also been restructured to offer more flexibility in defining the horizontal and vertical spatial coverage of the data sets. New sub-fields have been added, including: "Granule_Spatial_Representation", "Geometry", and "Orbit_Parameters".
Related_URL	Specifies links that contain information related to the data.	Related URL is now a required field. New sub-fields have also been added, including: "Protocol", "Title" and "Mime_Type".
Re-named/Re-structured/Removed Fields		
Entry_ID	The unique document identifier of the metadata record.	Entry ID now consists of two sub-fields: "Short_Name" is the short name associated with the collection, and "Version" is the version identifier of the data set.
DataSet_Citation	The citation information associated with the data set.	The Digital Object Identifier (DOI) sub-field has been updated to Persistent Identifier with the addition of a type field. This change was made to avoid confusion that the field was only to record DOIs.
Science_Keywords	The Earth science keywords that are representative of the data set.	A name change from "Parameters" to "Science_Keywords" was done to align the DIF schema with how Science Keywords are used elsewhere.
Ancillary_Keyword	Uncontrolled words or phrases used to further describe a data set.	A name change from "Keyword" to "Ancillary_Keyword" was done to clearly distinguish from "Science_Keywords" in the DIF schema.

Platform	The name of the platform used to acquire the data.	The Platform group has been updated to include Instruments/Sensors as sub-fields. New sub-fields have also been added, allowing additional information about platforms and associated instruments to be provided. Platform is now a required field.
Dataset_Progress	The production status of the data set regarding its completeness.	A name change from "Data_Set_Progress" to "Dataset_Progress" was done to better align the DIF schema with how this field is used elsewhere.
Personnel	The point of contact for more information about the data set or the metadata.	The Personnel field has been re-structured to capture two options: "Contact_Person" and "Contact_Group".
Organization	The data center, organization, or institution responsible for distributing the data.	The Data Center field has been renamed Organization. It has also been updated so users can supply more information about the different organizations and their roles in providing the data sets.
Sensor_Name (Instrument)	The name of the instrument used to acquire the data.	Sensor Name (Instrument) was moved to the Platform field.

Metadata_Creation	Specifies the date the metadata record was created.	"DIF_Creation_Date" was changed to "Metadata_Creation" in the DIF schema. It is now a sub-field of the Metadata Dates field.
Metadata_Last_Revision	Specifies the date the metadata record was last revised.	"Last_DIF_Revision_Date" was changed to "Metadata_Last_Revision" in the DIF schema. It is now a sub-field of the Metadata Dates field.
Metadata_Future_Review	Specifies a future date at which the DIF should be reviewed for accuracy of scientific or technical content.	"Future_DIF_Review_Date" was changed to "Metadata_Future_Review" in the DIF schema. It is now a sub-field of the Metadata Dates field.
Discipline	The science discipline associated with the data collection.	The Discipline field was removed from the DIF as the old GCMD Discipline data is no longer needed. Existing values are to be moved to Extended Metadata.