ESDS- RFC
for ECHO Metadata Standard

1 Status of this Memo

This memo provides information to the NASA Earth Science Data Systems (ESDS) community. This memo specifies an ESDS standard for Earth Science metadata. Distribution of this memo is unlimited.

2 Change Explanation

N/A

3 Copyright Notice

4 Abstract

This document defines the Earth Observing System (EOS) Clearinghouse (ECHO) metadata requirements and recommendations for ingesting science metadata into the ECHO system.

For each metadata type, the minimum metadata fields required to validate against the ECHO Ingest schema are outlined. In addition, a list of recommended metadata fields that should be included in data ingested into ECHO is provided. These additional fields will facilitate the searching and ordering of data by the Earth Science community.

The following three metadata constructs are utilized by the ECHO system:

- **Collection** - A grouping of science data that all come from the same source, such as a modeling group or institution. Collections have information that is common across all the granules they contain and a template for describing additional attributes not already part of the metadata model.
- **Granule** - The smallest aggregation of data that can be independently managed (described, inventoried, and retrieved). Granules have their own metadata model and support values associated with the additional attributes defined by the owning collection.
- **Browse** - An image which provides a high-level view of the associated granule or collection metadata item. Browse images are not spatially enabled, but are very useful during data discovery and cross-referencing to other granules or collections.

ECHO receives inventory metadata from ECS's Bulk Metadata Generator Tool (BMGT) as well as from non-ECS data providers who generate metadata files in the XML format that complies

with the published ECHO Metadata Model. Conformance to the model is verified by XML schemas.

ECHO's metadata model was derived from the ECS metadata model with some extensions. ECS data model in turn was developed in parallel to the FGDC model. Metadata conformant with ECS or generally with FGDC can be mapped to ECHO.

ECHO's metadata model is also being mapped to the core elements of the ISO 19115 standards for ingesting science metadata into the ECHO system and for representing query results produced by the ECHO system. A draft version of the query mapping is available. The final mappings of both will be complete within the 1st quarter of CY2010.

5 Table of Contents

6 Introduction

The ECHO 10.0 metadata requirements and recommendations outline the minimum metadata fields required to validate against the ECHO 10.0 Ingest schema for metadata Ingested into the ECHO system.

7 Our work

ECHO is a middleware solution which provides a Service Oriented Architecture (SOA) environment for the Earth Observing (EO) community. This environment serves both the providers and consumers of EO resources (data and services). ECHO provides common mechanisms for provider communities to publish their data and service offerings and other mechanisms for consumers to discover, understand, and access those resources.

By providing an SOA platform, ECHO allows the legacy providers of large amounts of data to focus on their contribution to the EO community to manage the availability of their information resources. These legacy providers can be freed from some of the secondary responsibilities of supporting discovery and usage of that data. ECHO works along with its Earth Science Data Partners to gather metadata representing each partner's data holdings in a process called "metadata ingest." This metadata is then made available through a published Application Program Interface (API) which exposes the necessary functionality for data discovery. ECHO also acts as an order broker between its Data Partners and end users.

ECHO also supports the diversified user community by providing common programmatic interfaces, based on standards, for all stakeholder communities. By providing these programmatic interfaces to a common infrastructure, ECHO enables the different stakeholder communities to build their own applications and user experiences that meet their user community needs, leveraging this common infrastructure. The services provided by the ECHO API are accessed through internally and externally developed client applications. ECHO's Client Partners work in cooperation with the ECHO team to develop efficient and specialized user interfaces which access the ECHO API.

The multi-organizational content of ECHO provides a valuable new service to a growing number of Earth science applications and interdisciplinary research efforts. ECHO streamlines access to digital data and materials and brokers orders and other services from Clients to Data Partners. ECHO provides tracking services for both the Provider and the Client.

7.1 Metadata Model

Chapter 4 of the ECHO Data Partner User’s Guide, version 10.6, January 2010, provides the specification for the ECHO Metadata Model.

http://www.esdswg.net/spg/rfc/esds-rfc-020/ECHO-10-Data-Partner-User-Guide-v-10.6.pdf

The full description of the ECHO Metadata Model including examples is referenced in the schema and documentation found on the ECHO Website.

(http://www.echo.nasa.gov/ingest/schemas/operations/docs/).
Information here is meant to be illustrative of the content of the ECHO Data Partner User's Guide, which is the sole specification document.

The metadata fields included in the ECHO Data Partner User's Guide are cited below and are designated as **required** or **recommended**.

- **Required Fields** – The metadata elements that must be present in order to pass the base XML Schema validation.
- **Recommended Fields** – The metadata elements that we recommend to facilitate searching and data usability by the science community.

The metadata name and description of each metadata field are listed in tabular form, along with the data type and relevant constraints, where applicable. Only the top-level elements of complex metadata elements are listed. Where this occurs, the requisite child element information is inherently required to correctly represent the parent metadata element.

COLLECTIONS - Required Elements

Name	Description
ShortName	This attribute identifies the short name associated with the collection. This is the official reference name used in identifying the contents of the data collection. All characters must be in upper case.
VersionId	This attribute specifies the version identifier of the data collection
InsertTime	This attribute specifies the insert date/time the collection entered the data provider's database. This date is provided by the data provider.
LastUpdate	This attribute specifies the most recent date/time an update occurred in the data provider's database. This date is provided by the data provider.
LongName	This attribute will identify the long name associated with the collection. This is the reference name used in describing the scientific contents of the data collection.
DataSetId	This attribute specifies a unique name for the collection. This information is computed by ECHO according to the data provider's policy.
Description	This attribute identifies the major emphasis of the content of the collection.
Orderable	This attribute indicates whether this collection is orderable.
Visible	This attribute indicates whether the collection is visible. Visibility is a basic access control mechanism that bypasses all ACL rules. If a collection is not visible, only users with the owning provider role will be able to see the item. No other users will be able to see the item—no matter what ACL permissions are in place. If group based permissions are needed, use the Restriction Flag field instead of visibility. Visibility is more commonly set at the collection level than the granule level. If a collection is not visible, none of the granules in the collection will be visible.

COLLECTIONS - Recommended Elements

Name	Description
ProcessingLevelId	The processing level class contains the level identifier and level description of the collection.
Price	This attribute specifies the price for ordering the collection
SpatialKeywords	This attribute specifies a word or phrase that serves to summarize the spatial regions covered by the collection. It may be repeated if several regions are covered. This often occurs when a collection is described as covering some large region, and several smaller sub regions within that region.
TemporalKeywords	This attribute specifies a word or phrase that serves to summarize the temporal characteristics referenced in the collection.
Temporal	This attribute contains records that describe the basis of the time system used for a specific collection.

Name	Description
Contact	This attribute contains the basic characteristics for a person or an organization type of contact. These contacts may provide information about a Collection, Delivered Algorithm Package, PGE, or Data Originator. System and user profile contact information is held elsewhere.
ScienceKeywords	This attribute holds a cross reference between collections and science keywords. The Science Keyword list is managed by the Global Master Change Directory (GCMD).
Platform	This attribute describes the relevant platforms associated with the acquisition of the collection or granule. Platform types include Spacecraft, Aircraft, Vessel, Buoy, Platform, Station, Network, Human etc.
Instrument	This entity registers the device used to measure or record data, including direct human observation. In cases where instruments have a single sensor or the instrument and sensor are used synonymously (e.g., AVHRR), both the Instrument and Sensor should be recorded. The Sensor information is represented by other entities.
Sensor	This entity holds the referential information for the collection source/sensor configuration including sensor parameter settings such as technique, etc.
Campaigns	This entity contains attributes describing the scientific endeavor(s) to which the collection is associated. Scientific endeavors include campaigns, projects, interdisciplinary science investigations, missions, field experiments, etc.
TwoDCoordinateSystem	This entity defines the two dimensional coordinate systems for the collection. The two dimensional coordinate system information is an alternative way to express spatial coverage. Granules in the collection that specify two dimensional coordinate data must conform to one of the systems defined by the collection.
OnlineAccessURL	This entity stores the online URL(s) for the granule, if there any. The URL either provides the site the user can obtain granule data or gives further instructions for obtaining the granule data.
OnlineResource	This entity records the documentation information of the collection including documentation type and documentation URL where applicable. These resources may include additional subsetting or processing services available for the granule.
AssociatedDIFs	This entity specifies the collection's Directory Interchange Format (DIF) identifier. The DIF identifier is used to uniquely identify a provider's data set. ECHO coordinates this list with the GCMD portal to facilitate the discovery of collections through GCMD. Providers may specify their own format, but the following is recommended: <ShortName>_V<Version_ID> (ex: MOD14_V005). Notice that the version id is zero padded to be three digits long.
Spatial	This entity contains the collection's spatial coverage information.
ArchiveCenter	Center where the collection is archived.
AdditionalAttributes	This entity stores the product specific attributes (i.e. attributes used to describe the unique characteristics of the collection which extend beyond those defined in this model).
AssociatedBrowseImages	List of browse images associated with the collection.

GRANULES - Required Elements

Name	Description

Name	Description
GranuleUR	The Universal Reference (UR) ID of the granule referred to by the data provider. This ID is unique per data provider.
InsertTime	This attribute is the date/time this granule entered the data provider's database.
LastUpdate	This attribute is the date/time the data provider last updated the granule information in the data provider's database
Collection	This attribute holds the reference information for a granule to a collection. The Granule references the collection either by collection short name and collection version or by collection dataset ID.
Orderable	This attribute indicates whether the granule is orderable.

GRANULES - Recommended Elements

Name	Description
DataGranule	This entity stores the basic descriptive characteristics associated with a granule.
PGEVersionClass	This entity stores basic descriptive characteristics related to the Program Generation Executable (PGE) associated with a granule.
Temporal	This entity contains records that describe the basis of the time system used for a specific collection.
Spatial	This entity contains the granule's spatial coverage information.
MeasuredParameters	This attribute specifies the names of the geophysical parameters expressed in the data as well as associated quality flags and quality status. The quality status contains measures of quality for the granule. The parameters used to set these measures are not preset and will be determined by the data producer. Each set of measures can occur many times either for the granule as a whole or for individual parameters. The quality flags contain the science, operational, and automatic quality flags that indicate the overall quality assurance levels of specific parameter values within a granule.
Platform	This attribute describes the relevant platforms associated with the acquisition of the collection or granule. Platform types include Spacecraft, Aircraft, Vessel, Buoy, Platform, Station, Network, Human, etc.
Instrument	This attribute registers the device used to measure or record data, including direct human observation. In cases where instruments have a single sensor or the instrument and sensor are used synonymously (e.g., AVHRR) both the Instrument and Sensor should be recorded. The Sensor information is represented by other entities.
Sensor	This attribute holds the referential information for the granule source/sensor configuration including sensor parameter settings such as technique, etc.
Campaigns	This entity contains attributes describing the scientific endeavor(s) to which the granules is associated. Scientific endeavors include campaigns, projects, interdisciplinary science investigations, missions, field experiments, etc.
DataFormat	This attribute contains the file format of the raw data (such as HDF) for this granule

Name	Description
TwoDCoordinateSystem	This entity stores the two dimensional coordinate system information for the granule. The two dimensional coordinate system information is an alternative way to express granule's spatial coverage based on a certain two dimensional coordinate system defined by the providers.
Price	This attribute specifies the price of the granule data when ordered.
OnlineAccessURL	This entity stores the online URL(s) for the granule, if any. The URL either provides the site the user can obtain granule data or gives further instructions for obtaining the granule data.
OnlineResource	This entity records the documentation information of the granule including documentation type and documentation URL where applicable. These resources may include additional subsetting or processing services available for the granule.
CloudCover	A percentage value indicating how much of the area of a granule (the ECS data unit) has been obscured by clouds. It is worth noting that there are many different measures of cloud cover within the ECS data holdings and that the cloud cover parameter that is represented in the DataPool is specific to Data Set.
AssociatedBrowseImages	The list of associated browse images to this granule
AdditionalAttributes	This entity stores the Product Specific Attributes with value a granule associates. The attribute name and type must exist in the parent collection for this granule.

BROWSE - Required Elements

Name	Description
ProviderBrowseId	The unique id for the browse image given by the provider. This should be unique per provider.
FileName*	The name of the delivered browse image
FileSize*	The size of the delivered browse image in bytes. A browse image delivered that is a different size than indicated will be rejected.
FileURL**	The URL for this browse image. Include file URL when a browse image is not hosted by ECHO, e.g. a browse image is hosted by the provider.

***** Providers must provide the browse file name and size if ECHO will be hosting the browse image.

****** Providers must provide the browse URL if the browse image will be hosted external to ECHO.

BROWSE - Recommended Elements

Name	Description
InsertTime	This attribute specifies the date/time this browse image was created in the data provider's database.
LastUpdate	This attribute specifies the last time the browse image was modified in the provider's database.

7.1.1 Meeting User Needs - How has the ECHO model evolved to meet user needs?

- Transitioned responsibility of ECHO's definition management from the V0 definition server to GCMD Dataset Definitions for the purpose of displaying data set definitions.
- Adopted peer-reviewed GCMD Keyword Vocabulary for Discipline Topic Keywords in the ECHO schema
- Adjusted model to allow Data Providers to provide urls to their externally hosted Browse files in addition to having ECHO host the Browse file
- Providing web-applications PUMP (Provider Utility Management Program) and EIAT (ECHO Ingest Accounting Tool) to manage data and orders in ECHO and to track live and completed Ingest jobs.

7.2 Tips for preparing Metadata

There are separate schemas that govern the XML format for the various metadata types, available from the ECHO website at http://www.echo.nasa.gov

Click: Data Partners → Data Development

There are also examples of minimum collection and granule metadata XML requirements located on the ECHO website at http://www.echo.nasa.gov

Click: Data Partners → Getting Started

7.3 Collaborations (i.e. Who's using the ECHO Metadata Model)

7.3.1 Data Partners

The following NASA EOSDIS data centers are publishing their data holdings in ECHO using the ECHO Metadata Model:

ASF - The Alaska Satellite Facility (ASF) is under contract to acquire, process, archive, and distribute satellite Synthetic Aperture Radar (SAR) data for research communities:

http://www.asf.alaska.edu/

GES DAAC - GSFC Earth Sciences, Distributed Active Archive Center for upper atmosphere, atmospheric dynamics, global land biosphere, global precipitation and ocean color:

http://daac.gsfc.nasa.gov/

GHRC - The Global Hydrology Resource Center (GHRC) provides both historical and current Earth science data, information, and products from satellite, airborne, and surface-based instruments:

http://ghrc.msfc.nasa.gov/

LAADS – Level 1 and Atmosphere Archive and Distribution System for MODIS Level 1 and atmosphere: http://ladsweb.nascom.nasa.gov/

LaRC DAAC - Langley Research Center, Distributed Active Archive Center for radiation budget, clouds, aerosols and tropospheric chemistry:

http://eosweb.larc.nasa.gov/

LP DAAC - Land Processes, Distributed Active Archive Center for land process data: http://edcdaac.usgs.gov/main.html

NSIDC DAAC - National Snow and Ice Data Center, Distributed Active Archive Center for snow and ice, cryosphere and climate:

http://nsidc.org/daac/

ORNL DAAC - Oak Ridge National Laboratory, Distributed Active Archive Center for biogeochemical dynamics and ecological data for studying environmental processes: http://www.daac.ornl.gov/

PO DAAC – Physical Oceanography, Distributed Active Archive Center for global oceanographic: http://podaac.jpl.nasa.gov/

SEDAC - Socioeconomic Data and Applications Center for population, sustainability, geospatial data, multilateral environmental agreements:

http://sedac.ciesin.columbia.edu/

USGS_EROS - The Landsat Project is a joint initiative of the U.S. Geological Survey (USGS) and NASA designed to gather Earth resource data from space. The 35-year record of images provides a unique resource for people who work in agriculture, geology, forestry, regional planning, education, mapping, and global change research:

http://landsat.usgs.gov/index.ph

7.3.2 Client Partners

The following operational ECHO clients are accessing ECHO data holdings for search and/or order using the ECHO Metadata Model:

WIST - The Warehouse Inventory Search Tool is a web-based client for searching and ordering earth science data from various NASA and affiliated centers. Users can submit cross-discipline queries using spatial and temporal criteria, examine search results for relevancy using built-in tools, and submit orders via ECHO to the appropriate data provider.

https://wist.echo.nasa.gov

SNOWI - The Search 'N Order Web Interface tool is a web-based client used to search a select set of NSIDC data holdings. It is intended as a lightweight tool for quick access to data for its user community.

http://nsidc.org/data/snowi/index.html

ESA - The European Space Agency's latest client allows access into the ECHO holdings. The new EOLI Web Client provides access to ESA catalogues of EO products via a standard web browser, as well as to the catalogues of other data providers (for example DLR and NASA ECHO) and is part of ESA's eoPortal.

http://catalogues.eoportal.org/eoli.html

8 References

Normative References

1. ECHO Data Partner User Guide version 10.6, January 2010. EOS Clearinghouse (ECHO). National Aeronautics and Space Administration.

http://www.esdswg.net/spg/rfc/esds-rfc-020/ECHO-10-Data-Partner-User-Guide-v-10.6.pdf

2. ECHO Metadata Schema and DTDs, 2008. EOS Clearinghouse (ECHO). National Aeronautics and Space Administration.

   http://www.echo.nasa.gov/ingest/schemas/operations/docs/index.html

Informative References

1. ECHO Main Website: http://www.echo.nasa.gov/index.html

9 Authors' Address

Andrew Mitchell NASA GSFC Greenbelt Road Greenbelt Maryland 20771 USA Tel: 301-614-5189 fax 301-614-5270, email: Andrew.E.Mitchell@nasa.gov

10 Appendix A – Glossary of Acronyms

Acronym	Description
API	Application Programming Interface
ASF DAAC	Alaska Satellite Facility DAAC
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
COTS	Commercial Off The Shelf
DAAC	Distributed Active Archive Center
DB	DataBase
DTD	Document Type Definition
ECHO	EOS Clearinghouse
ECS	EOSDIS Core System

EDC	EROS Data Center
EMD	EOSDIS Maintenance and Development
EOS	Earth Observing System
EOSDIS	EOS Data and Information System
EROS	Earth Resources Observation Systems
ESDIS	Earth Science Data and Information System
ESIP	Earth Science Information Partner
ETC	ECHO Technical Committee
FTP	File Transfer Protocol
GCMD	Global Change Master Directory
GES DAAC	GSFC Earth Sciences DAAC
GHRC	Global Hydrology Resource Center
GIS	Geographic Information System
GML	Geography Markup Language
GMT	Greenwich Mean Time
GSFC	Goddard Space Flight Center
GUI	Graphical User Interface
GUID	Globally Unique Identifier
IIMS	Independent Information Management Subsystem
LAADS	Level 1 and Atmosphere Archive and Distribution System
LP DAAC	Land Processes DAAC
MISR	Multiangle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NSIDC DAAC	National Snow and Ice Data Center DAAC
ODL	Object Description Language
OGC	OpenGIS Consortium
ORNL DAAC	Oak Ridge National Laboratory DAAC
PGE	Product Generation Executives
PO.DAAC	Physical Oceanography DAAC

PSA	Product Specific Attribute
PUMP	Provider User Management Program
QA	Quality Assurance
SEDAC	Socioeconomic Data and Applications Center
SOAP	Simple Object Access Protocol
SSC	Stennis Space Center
SSL	Secure Sockets Layer
UDDI	Universal Description, Discovery, and Integration
UI	User Interface
UR	Universal Reference
UR	Uniform Resource Identifier
URL	Uniform Resource Locator
UTC	Universal Time, Coordinated (also called GMT/UTC)
WIST	Warehouse Inventory Search Tool
WGS	World Geodetic System
WRS	Worldwide Reference System
WSDL	Web Services Description Language
XML	eXtensible Markup Language
XSLT	eXtensible Style Language Transformation