

Metadata in EOSDIS

V0, ECS, FGDC, ECHO and ISO

Richard Ullman

ESDIS Project

HDF & HDF-EOS Workshop

September 24, 2003



Metadata in EOSDIS

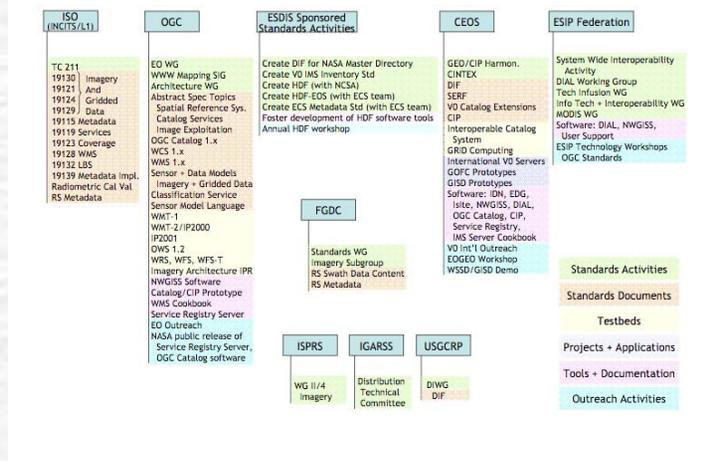
V0, ECS, FGDC, ECHO and ISO

Richard Ullman, ESDIS Project

A Three-level Data Model

- **Collection** - A collection is a set of files or data granules. Each member of a collection is similar to other members because it contains the the same or similar kind of content, source and format. Collections are sometimes called data sets or data series. Within ECS a collection is called an Earth Science Data Type (ESDT)
- **Granule** - a particular member of a collection. Often a file, sometimes a group of files. Within EOSDIS, a granule is the smallest managed unit of data in the archive.
- **Sub-granule** - services on granules may access to parts within a granule such as geographic subsets or parameter selections. Services can deliver smaller units to end users.

ESDIS Standards Involvement



Metadata Lessons

- Necessary metadata is more varied and complex than first appears.
 - Data producers require metadata model to be extendable to accommodate specific operational parameters.
- Metadata "population" and "valids"
 - Metadata must be reliably populated to be useful.
 - It is difficult to "correct" metadata after it is entered into inventory.
 - Tools that provide sufficient feedback markedly improve metadata population and accuracy (e.g. EDG EVP).
- For search:
 - Vast majority of users use only dataset name, time and location information.
 - Some communities have a small number of additional favorite attributes (e.g. Land imagery users often query on cloud cover)
 - "Advanced" ordering systems use more attributes. (e.g. EOSDIS Data Validation User Interface for MODIS Land Validation Users)

Metadata Standards

- Established practice is at least as important as the myriad of national and international standards and working groups.
- Standards for metadata and interoperability have evolved in tandem with EOSDIS evolution.
- Standards will continue to evolve over the foreseeable future.

V0 Interoperability Protocol

- The Version Zero protocol is a search and order query protocol designed to link heterogeneous data archives into a single search.
- V0 links NASA DAACs and several international partners.
 - 7 "core" attributes + geographic and temporal parameters. Many hundreds of "extended attributes".
 - Centralized collection metadata.
 - Distributed granule metadata
 - Search is distributed. Full results depend on "weakest link"

ECS (EOSDIS Core System)

- ECS includes a full archive management implementation including ingest/delete, inventory management, user accounts, order processing, etc.
- Inventory is based on hierarchical collection/granule data model that includes approximately 300 core attributes and many hundreds of "additional attributes".
- Access to inventory is through V0 protocol, "data pools", DAAC specific means.

GCMD (Global Change Master directory)

- GCMD is NASA's minimum interoperability standard for NASA's Earth Observation Data.
 - Lists collection metadata links to data center.
 - GCMD metadata directly maps to the FGDC as well as many other standards
 - <http://gcmd.gsfc.nasa.gov/Aboutus/standards/>
 - About 90 attributes.
- GCMD is the NASA node for:
 - the FGDC Clearinghouse .
 - e-gov Geospatial One-Stop
 - Committee on Earth Observation Satellites International Directory Network (CEOS IDN)

FGDC (Federal Geographic Data Committee)

- NASA has spent considerable effort in harmonization of NASA systems with FGDC.
- The most relevant standards are:
 - FGDC-STD-001 Content Standard for Geospatial Metadata - GCMD, ECS, and ECHO are compliant
 - FGDC-STD-009 Standard for Remote Sensing Swath Data (Content) - HDF-EOS Swath content is basis for this standard.
 - FGDC-STD-012 Content Standard for Digital Geospatial Metadata - Extensions for Remote Sensing Metadata - ECS data model is basis for this standard

ISO/ CEOS/ OGC

- NASA projects, particularly the Earth Science Data Information Systems project and the Geospatial Interoperability Program, have been proactive participants in both ISO and CEOS.
- Acronyms
 - International Standards Organization Technical Committee 211: Geographic Information/Geomatics (ISO TC211)
 - Committee on Earth Observation Satellites Working Group on Information Systems and Services (CEOS WGISS)
 - OpenGIS Consortium (OGC),

HDF & HDF-EOS Workshop, September 24, 2003



A Three-level Data Model

- Collection - A collection is a set of files or data granules. Each member of a collection is similar to other members because it contains the the same or similar kind of content, source and format. Collections are sometimes called data sets or data series. Within ECS a collection is called an Earth Science Data Type (ESDT)
- Granule - a particular member of a collection. Often a file, sometimes a group of files. Within EOSDIS, a granule is the smallest managed unit of data in the archive.
- Sub-granule - services on granules may access to parts within a granule such as geographic subsets or parameter selections. Services can deliver smaller units to end users.



Metadata Standards

- Established practice is at least as important as the myriad of national and international standards and working groups.
- Standards for metadata and interoperability have evolved in tandem with EOSDIS evolution.
- Standards will continue to evolve over the foreseeable future.



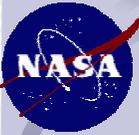
V0 Interoperability Protocol

- ☞ The Version Zero protocol is a search and order query protocol designed to link heterogeneous data archives into a single search.
- ☞ V0 links NASA DAACs and several international partners.
 - 7 “core” attributes + geographic and temporal parameters. Many hundreds of “extended attributes”.
 - Centralized collection metadata.
 - Distributed granule metadata
 - Search is distributed. Full results depend on “weakest link”



ECS

- ECS includes a full archive management implementation including ingest/delete, inventory management, user accounts, order processing, etc.
- Inventory is based on hierarchical collection/granule data model that includes approximately 300 core attributes and many hundreds of “additional attributes”.
- Access to inventory is through V0 protocol, “data pools”, DAAC specific means.



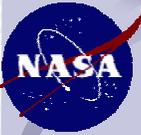
HDF-EOS

- ECS Science Metadata is stored at granule creation in the HDF file as a text attribute "coremetadata.n"
- The metadata is encoded using ODL (Object Description Language).



GCMD (Global Change Master directory)

- ☛ GCMD is NASA's minimum interoperability standard for NASA's Earth Observation Data.
 - Lists collection metadata links to data center.
 - GCMD metadata directly maps to the FGDC as well as many other standards
 - <http://gcmd.gsfc.nasa.gov/Aboutus/standards/>
 - About 90 attributes.
- ☛ GCMD is the NASA node for:
 - the FGDC Clearinghouse .
 - e-gov Geospatial One-Stop
 - Committee on Earth Observation Satellites International Directory Network (CEOS IDN)



Federal Geographic Data Committee (FGDC)

- ☞ NASA has spent considerable effort in harmonization of NASA systems with FGDC.
- ☞ The most relevant standards are:
 - FGDC-STD-001 Content Standard for Geospatial Metadata - GCMD, ECS, and ECHO are compliant
 - FGDC-STD-009 Standard for Remote Sensing Swath Data (Content) - HDF-EOS Swath content is basis for this standard.
 - FGDC-STD-012 Content Standard for Digital Geospatial Metadata - Extensions for Remote Sensing Metadata - ECS data model is basis for this standard



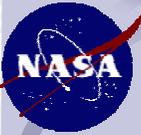
ISO/ CEOS/ OGC

- ☛ NASA projects, particularly the Earth Science Data Information Systems project and the Geospatial Interoperability Program, have been proactive participants in both ISO and CEOS.
- ☛ Acronyms
 - International Standards Organization Technical Committee 211: Geographic Information/Geomatics (ISO TC211)
 - Committee on Earth Observation Satellites Working Group on Information Systems and Services (CEOS WGISS)
 - OpenGIS Consortium (OGC),



Metadata Lessons

- ☞ Necessary metadata is more varied and complex than first appears.
 - Data producers require metadata model to be extendable to accommodate specific operational parameters.
- ☞ Metadata “population” and “valids”
 - Metadata must be reliably populated to be useful.
 - It is difficult to “correct” metadata after it is entered into inventory.
 - Tools that provide sufficient feedback markedly improve metadata population and accuracy (e.g. EDG EVP).



Metadata Lessons

For search:

- Vast majority of users use only dataset name, time and location information.
- Some communities have a small number of additional favorite attributes (e.g. Land imagery users often query on cloud cover)
- “Advanced” ordering systems use more attributes. (e.g. EOSDIS Data Validation User Interface for MODIS Land Validation Users)



**ISO
(INCITS/L1)**

- TC 211
- 19130 Imagery
- 19121 And
- 19124 Gridded
- 19129 Data
- 19115 Metadata
- 19119 Services
- 19123 Coverage
- 19128 WMS
- 19132 LBS
- 19139 Metadata Impl.
- Radiometric Cal Val
- RS Metadata

OGC

- EO WG
- WWW Mapping SIG
- Architecture WG
- Abstract Spec Topics
- Spatial Reference Sys.
- Catalog Services
- Image Exploitation
- OGC Catalog 1.x
- WCS 1.x
- WMS 1.x
- Sensor + Data Models
- Imagery + Gridded Data
- Classification Service
- Sensor Model Language
- WMT-1
- WMT-2/IP2000
- IP2001
- OWS 1.2
- WRS, WFS, WFS-T
- Imagery Architecture IPR
- NWGISS Software
- Catalog/CIP Prototype
- WMS Cookbook
- Service Registry Server
- EO Outreach
- NASA public release of Service Registry Server, OGC Catalog software

ESDIS Sponsored Standards Activities

- Create DIF for NASA Master Directory
- Create V0 IMS Inventory Std
- Create HDF (with NCSA)
- Create HDF-EOS (with ECS team)
- Create ECS Metadata Std (with ECS team)
- Foster development of HDF software tools
- Annual HDF workshop

FGDC

- Standards WG
- Imagery Subgroup
- RS Swath Data Content
- RS Metadata

ISPRS

- WG II/4 Imagery

IGARSS

- Distribution Technical Committee

USGCRP

- DIWG
- DIF

CEOS

- GEO/CIP Harmon.
- CINTEX
- DIF
- SERF
- V0 Catalog Extensions
- CIP
- Interoperable Catalog System
- GRID Computing
- International V0 Servers
- GOFC Prototypes
- GISD Prototypes
- Software: IDN, EDG, Isite, NWGISS, DIAL, OGC Catalog, CIP, Service Registry, IMS Server Cookbook
- V0 Int'l Outreach
- EOGEO Workshop
- WSSD/GISD Demo

ESIP Federation

- System Wide Interoperability Activity
- DIAL Working Group
- Tech Infusion WG
- Info Tech + Interoperability WG
- MODIS WG
- Software: DIAL, NWGISS, User Support
- ESIP Technology Workshops
- OGC Standards

- Standards Activities
- Standards Documents
- Testbeds
- Projects + Applications
- Tools + Documentation
- Outreach Activities