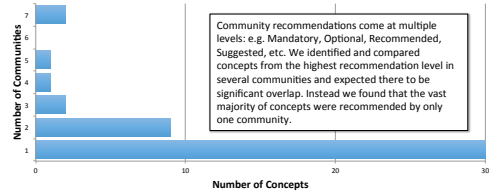


Big Earth Data Initiative (BEDI) Metadata Improvement: Case Studies (IN21C-1699)

Ted Habermann, John Kozimor, Sean Gordon and Lindsay Powers, The HDF Group (thabermann@hdfgroup.org), John Farley, Raytheon

What metadata are important?

It depends on the use case: discovery, access, use, or understanding and on what is being described (project, data collection site, dataset). There are existing metadata recommendations for all these use cases and resource types. They are created by organizations and communities to address specific needs.



Community Recommendation

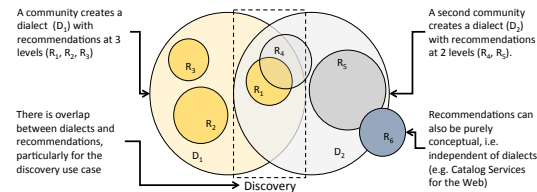
Concept	ACDD Highly Recommended	CSW Core Queryables	DIF Required	ECHO Mandatory	FGDC Mandatory	ISO 15115 Mandatory	UIMM Common Required
Abstract	X	X	X	X	X	X	X
Resource Title	X	X	X	X	X	X	X
Theme Keyword	X	X	X	X	X	X	X
Modified Date			X			X	X
Bounding Box		X			X	X	
Distribution Contact			X				X
Metadata Contact					X	X	
Metadata Identifier		X	X				
Metadata Standard Citation			X		X		
Metadata Standard Version			X		X		
Ordering Instructions				X	X		
Publication Date			X	X			
Resource Access Constraints			X	X			
Temporal Extent				X			X
Topic Category			X			X	
Unique Concepts	Association, Coordinate Reference System, Resource Identifier, Resource Format, Resource Type	Resource Identifier, Long Name, Resource Version	Distribution Liability, Originating Organization, Purpose, Resource Status, Resource Update Frequency, Resource Use Constraints, Turnaround	Browse File Name, Browse URL, Future Metadata Review Date, Instrument, Instrument Keyword, Instrument Short Name, Metadata Deletion Date, Platform, Platform Keyword, Platform Short Name, Processor, Resource Contact, Resource on-line Link, Sensor Short Name, Spatial Extent			

Metadata Dialects

Many communities use the term "standard" when they describe their metadata and, as a result, there are many existing "standards". This approach focuses attention on differences between communities. We use the term "dialect" to focus attention on common concepts and goals.

Recommendations and Dialects:

Recommendations reflect community experiences and documentation needs. Communities have common documentation needs, so recommendations overlap, particularly for the discovery use case. Sharing recommendations is an important mechanism for sharing those experiences and community knowledge.

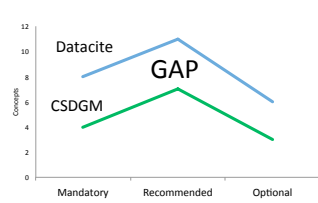


There is overlap between dialects and recommendations, particularly for the discovery use case

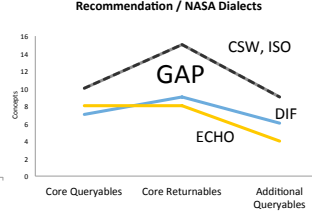
Metadata Recommendations and Dialects

Metadata recommendations change as new communities and needs emerge. Metadata management tools are driven by dialects. Changing those tools and training people are difficult, so adoption of new dialects is relatively slow. This leads to gaps between existing organizational capabilities (dialects) and recommendations.

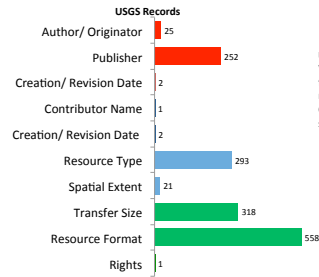
DataCite Recommendation / CSDGM Dialect



Catalog Services for the Web Recommendation / NASA Dialects



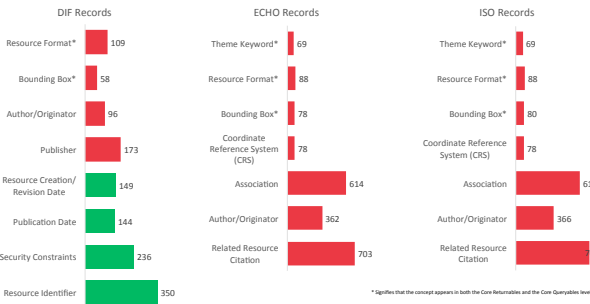
Evaluating Completeness of Metadata Collections



U.S.G.S. Science Base / DataCite
We examined completeness of 617 CSDGM records from 12 collections in the U.S.G.S. Science Base with respect to the Mandatory, Recommended, and Optional DataCite recommendations. The Figure shows the number of records missing fields.

NASA Climate Data Initiative

We examined completeness of metadata for the NASA Climate Data initiative in three dialects (DIF, ECHO, and ISO) with respect to the Catalog Services for the Web Core Queryables, Core Returnables, Additional Queryables. The Figure shows the number of records missing fields.



* Specifies that the concept appears in both the Core Returnables and the Core Queryables level

People and Organizations in ESDIS Metadata Collections

Connecting users to people and organizations that can help them access, use, and understand data is an important metadata role. The ESDIS Common Metadata Repository (CMR) includes a responsibility element that defines responsibilities related to data resources. The UIMM Common Recommendation defines five responsibilities: Metadata Contact, Resource Author / Originator, Point of Contact, Distributor, and Processor. Understanding usage, completeness and consistency of Responsibilities and Parties in ESDIS metadata is an important first step towards providing consistent and complete services to users of those data.

We examined 2158 metadata records from 15 collections. These tables show the number of elements / record for Authors and Processors. Red cells indicate missing elements.

Author / Originator:

Path Elements	ADP	ODIN	ORNL	SESTRA	TRISA	LANDROVER	LANC	LANC_ARE	SPINAL_ARE	NESTOR	NESTOR_ARE	ORBITER	PODAAC	SEDA	USRA_ARE
administrativeArea															
city															
country															
deliveryPoint															
electronicMailAddress															
postalCode															
contactInstructions															
hoursOfService															
phone/gnd/facsimile															
phone/gnd/voice															
individualName															
organizationName															
positionName															
Role (codeList-value)															

Processor

Path Elements	ADP	ODIN	ORNL	SESTRA	TRISA	LANDROVER	LANC	LANC_ARE	SPINAL_ARE	NESTOR	NESTOR_ARE	ORBITER	PODAAC	SEDA	USRA_ARE
organisationName															
Role (codeList-value)															

Contact Information / Email

Contact information is critical. This table identifies people and organizations in the metadata that include email addresses.

Path Elements	ADP	ODIN	ORNL	SESTRA	TRISA	LANDROVER	LANC	LANC_ARE	SPINAL_ARE	NESTOR	NESTOR_ARE	ORBITER	PODAAC	SEDA	USRA_ARE
Metadata Contact															
Distribution Contact															
Resource Author / Originator															
Point of Contact															
Processor															

How do dialects evolve?

When communities need metadata concepts that are not included in their native dialects, the dialects can be extended to include those concepts.

We examined 12 metadata collections (A-L) from the U.S.G.S. Science Base. Several of these have extended the CSDGM dialect to include citations to processing method descriptions. The extension includes 15 elements shown as rows in this Table. The Table shows the number of these elements / record in each collection.

Two groups (A and B) have complete descriptions of these citations for some records. Other groups have adopted several elements of the extension, most commonly method type (methtype) and description (methdesc).

Record Count	26	37	13	7	2	10	50	1	2	50	9	50
Collection	A	B	C	D	E	F	G	H	I	J	K	L
Element Name												
origin												
pubdate												
title												
geoform												
publish												
pubplace												
onlink												
issue												
sername												
methtype												
methdesc												
otherct												
edition												
methkey												
methkt												

Collections A-G have adopted some elements of the processing methods citation extension.

Collections H-L have not adopted the processing methods citation extension.



<http://www.nasa.gov>



This work was supported by NASA/GSFC under Raytheon Co. contract number NNG05SH293C
NASA Official: Jeff Walter (jeff.walter@nasa.gov)