

People and Organizations in CMR Metadata

Interim Collection Comparison Report for Big Earth Data Initiative

Metadata Source: NASA Common Metadata Repository

Metadata Dialect: ISO 19115-2

Evaluation Target: People and Organizations

Ted Habermann – The HDF Group

Metadata serves an essential function in connecting users to people and organizations to help them access, use, and understand data. The ESDIS Common Metadata Repository (CMR) recognizes the importance of this type of information and includes two related elements in the Unified Metadata Model (UMM) Common Profile: Responsibility and Party. As elements in the Common Profile, these elements are included in all other UMM Profiles.

The Responsibility element broadly defines responsibilities related to data resources using the position of the element in the metadata model hierarchy. The UMM-Common Profile defines five responsibilities: Metadata Contact, Resource Author / Originator, Point of Contact, Distributor, and Processor. Each of these responsibilities can have multiple people or organizations (termed *parties*) associated with it. A RoleCode that is chosen from the standard ISO Codelist describes details of the roles of those parties.

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. This report provides an initial assessment of these characteristics.

We examined 4180 metadata records from 17 collections extracted from the CMR during March 2015 and 2126 metadata records from 15 collections extracted from the CMR during October 2015 ([Collections Examined](#)). For each recommended responsibility we provide a table that gives the average number of occurrences for elements of the associated party in each of these collections. A value of 1 or more typically (although not necessarily) indicates that the element is included one or more times in each record in a collection. A value < 1.0 is typically the percentage of records in a collection that include the metadata element. Empty cells indicate values of 0 – the element is completely missing from the collection.

Contact Information in the CMR

Metadata Contact:

ISO path: /qmi:MI Metadata/qmd:contact/qmd:CI ResponsibleParty

The Metadata Contact responsibility gives the party that is responsible for creating and maintaining the metadata. The xPath of the source of this information in the DIF and ECHO dialects is given in Table 1.

Table 1. Sources for Metadata Contact information

Dialect	Source
DIF	/DIF/Personnel[Role="DIF AUTHOR"]
ECHO	1. /*/ArchiveCenter 2. /*/Contacts/Contact[contains(Role,'DIF AUTHOR')]

This table shows the frequency of occurrence of elements of the Metadata Contact Responsibility / Parties for 25 collections in the CMR. The data indicates that all of the collections include the name of the organization that is responsible for the metadata (gmd:organisationName) and a roleCode (gmd:role/gmd:CI RoleCode) for the organization. In general, no contact information is provided for these organizations with the exception of CDDIS.

This difference reflects the source of the information in ECHO. Metadata Contacts that originate as ArchiveCenter have only the organisationName while those that originate as contacts generally have more complete information.

Table 2. Occurrences of Metadata Contact Responsibility

[illegible]

Hours	gmd:identificationInfo/gmd:citation/gmd: citedResponsibleParty/gmd:contactInfo/gmd: hoursOfService								. 12				.71					
Fax	gmd:identificationInfo/gmd:citation/gmd: citedResponsibleParty/gmd:contactInfo/gmd:phone /gmd:facsimile								. 00				.02					
Phone	gmd:identificationInfo/gmd:citation/gmd: citedResponsibleParty/gmd:contactInfo/gmd:phone /gmd:voice								. 20			.82	.89					
Individual Name	gmd:identificationInfo/gmd:citation/gmd: citedResponsibleParty/gmd:individualName								. 16				1.17					
Organizational Name	gmd:identificationInfo/gmd:citation/gmd: citedResponsibleParty/gmd:organisationName											.82	.32					
Position Name	gmd:identificationInfo/gmd:citation/gmd: citedResponsibleParty/gmd:positionName								. 16				.57					
Role Code	gmd:identificationInfo/gmd:citation/gmd: citedResponsibleParty/gmd:role/gmd:C_I_RoleCode								. 16			.82	1.49					

Organizational Name	gmd:identificationInfo/gmd:pointOfContact/gmd:organisationName	1.00	3.86	1.00	3.71	3.39	2.02	4.00	3.90	3.56	1.00	3.07	1.75	1.00	2.00	1.05	4.20	1.00	5.38	1.00	.80	1.00	1.00	1.16	1.00	3.82
Role Code	gmd:identificationInfo/gmd:pointOfContact/gmd:role/gmd:CI_RoleCode	1.00	7.82	1.21	6.36	6.35	5.32	7.00	6.32	7.10	1.00	7.45	4.11	1.01	4.00	1.13	7.19	1.09	10.69	1.00	.80	1.00	1.00	1.39	1.00	6.51

Feb 2016 Version

Distribution Contact:

ISO Path: /gmi:MI_Metadata/gmd:distributionInfo/gmd:MD_Distribution/gmd:distributor/gmd:MD_Distributor

/gmd:distributorContact/gmd:CI_ResponsibleParty/

The Distribution Contact Responsibility contains the party responsible for answering questions about the distribution of the dataset. This is typically the Archive Center for the dataset. The xPath of the source of this information in the DIF and ECHO dialects is given in Table 7.

Table 7. Sources for Distribution Contact Responsibility information.

Dialect	Source
DIF	1. /DIF/Data_Center/Data_Center_Name
ECHO	1. /*/ArchiveCenter 2. /*/Contacts/Contact[contains(Role,'Archive')] 3. /*/Contacts/Contact[contains(Role,'DATA CENTER CONTACT')] 4. /*/Contacts/Contact[contains(Role,'Distributor')] 5. /*/Contacts/Contact[contains(Role,'User Services')] 6. /*/Contacts/Contact[contains(Role,'GHRC USER SERVICES')] 7. /*/Contacts/Contact[contains(Role,'ORNL DAAC User Services')]

Table 8 shows the frequency of occurrence of elements of the Distribution Contact Responsibility / Parties for 25 collections in the CMR.

Table 8. Occurrences of the Distributor Contact Responsibility

	Number of Records	161	2559	38	103	23	1044	19	340	130	6	154	406	606	1	285	5488	223	784	132	5	1216	603	202	11	130
Element	Element Path	A SF	AU _A ADC	C D DIS	E SA	EU ME TSAT	GE S_ DISC	IS RO	J A XA	L A A DS	LAN CEA MSR2	LAN CEM ODIS	L A RC	LAR C_A SDC	LM _FI RMS	LPD AAC _ECS	NO AA_ NCEI	NSI DC_ ECS	NS ID CV0	OB _D AAC	O MI N RT	ORN L_D AAC	PO DA AC	S E D AC	USG S_E ROS	US GS _LTA
Administrative Area	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:address/gmd:administrativeArea			.89									.36	1.00		.96		.91				1.00		.92	1.00	
City	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:address/gmd:city			.89									.36	1.00		.96		.91				1.00		.92	1.00	
Country	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:address/gmd:country			.89									.36	1.00		.96		.91				1.00		.92	1.00	
Delivery Point	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:address/gmd:deliveryPoint			.89									.36	1.00		.96		.91				1.00		.92	1.00	
Email	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:address/gmd:electronicMailAddress	1.00		.89						1.00			.36	1.00		.96		.91		1.00		1.00	1.09	.92	1.00	
Postal Code	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:address/gmd:postalCode			.89									.36	1.00		.96		.91				1.00		.92	1.00	
URL	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:onlineResource/gmd:linkage/gmd:URL		1.01		1.04	1.00		1.00	1.01	.72		.50					1.34		1.63					.08		1.02
Protocol	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:onlineResource/gmd:protocol		1.01		1.04	1.00		1.00	1.01	.72		.50					1.34		1.63					.08		1.02

Phone	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:phone/gmd:voice	1.00		1.79									.36	1.00		.96		.91				2.00	1.09	1.84	4.00		
Individual Name	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:individualName			.89									.11					.01				1.00	1.09	.92			
Organization Name	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:organisationName		1.01	.89	1.04	1.00			1.00	1.01	.72		.50	.25	1.99		.96	1.35	.91	1.63	1.00	.80		1.09	1.00	1.00	1.02
Role Code	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:role/gmd:CI_RoleCode	1.00	1.01	.89	1.04	1.00			1.00	1.01	.72	1.00	.50	.36	1.99		.96	1.35	.91	1.63	1.00	.80	1.00	1.09	1.00	1.00	1.02
Contact Instructions	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:contactInstructions												.36				.96		.91						1.00		
Hours	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:hoursOfService												.36				.96		.91					.92	1.00		
Fax	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:contactInfo/gmd:phone/gmd:facsimile												.36				.96		.91								
Position Name	gmd:distributionInfo/gmd:distributor/gmd:distributorContact/gmd:positionName												.11					.01					1.09				

Feb 2016 Version

Processor

ISO path: /gmi:MI_Metadata/gmd:dataQualityInfo/gmd:DQ_DataQuality/gmd:lineage/gmd:LI_Lineage

/gmd:processStep/gmd:LI_ProcessStep/gmd:processor

The Processor Responsibility gives the party that is responsible for processing the dataset. It is included in the lineage metadata. The xPath of the source of this information in the DIF and ECHO dialects is given in Table 9.

Table 9. Sources for Processor Responsibility information.

Dialect	Source
DIF	1. TBD
ECHO	1. /*/ProcessingCenter

Table 10 shows the frequency of occurrence of elements of the Processor Responsibility / Parties for 25 collections in the CMR.

Table 10. Occurrences of Processor Responsibility

	Number of Records	161	2559	38	103	23	1044	19	340	130	6	154	406	606	1	285	5488	223	784	132	5	1216	603	202	11	130
Element	Element Path	ASF	AU_AADC	CD DIS	ESA	EUTSAT	GE_SDISC	ISRO	JAXA	LANDMSR2	LANCER	LANCER	LARC	LARSDC	LM_FIRMS	LPD_AAC_ECS	NOAA_NCEI	NSIDC_ECS	NSIDC_CV0	OB_DAAC	OMI_NRT	ORNL_DAAC	PODAAC	SEDA	USGS_EROS	USGS_LTA
Organization Name	gmd:dataQualityInfo/gmd:lineage/gmd:processStep/gmd:processor/gmd:organisationName	.98							.72		.50	.25			.96		.91				1.00	1.00	.92	1.00		
Role Code	gmd:dataQualityInfo/gmd:lineage/gmd:processStep/gmd:processor/gmd:role/gmd:CI_RoleCode	.98							.72	.50	.25				.96		.91				1.00	1.00	.92	1.00		
URL	gmd:dataQualityInfo/gmd:lineage/gmd:processStep/gmd:processor/gmd:contactInfo/gmd:onlineResource/gmd:linkage/gmd:URL								.70	.50																
Protocol	gmd:dataQualityInfo/gmd:lineage/gmd:processStep/gmd:processor/gmd:contactInfo/gmd:onlineResource/gmd:protocol								.70	.50																

Feb 2016 Version

E-Mail Addresses

The data shown above clearly indicates that the completeness of contact information varies significantly across collections and responsibilities. The standards all include extensive physical contact information, e.g. cities, addresses, and postal codes. This reflects the prevalence of physical mail delivery when these standards were created. Now electronic delivery dominates, so e-mail addresses are more likely to be helpful than physical addresses. We examined the occurrence of e-mail addresses for responsibilities in all 15 collections. Table 12 gives the results of this analysis.

Table 12. Completeness of contact email addresses

	161	2559	38	103	23	1044	19	340	130	6	154	406	606	1	285	5488	223	784	132	5	1216	603	202	11	130
Element	A SF	AU_A ADC	CD DIS	E SA	EUME TSAT	GES_DISC	IS RO	JA XA	LA ADS	LANCE AMSR2	LANCE MODIS	LA RC	LARC_ASDC	LM_FI RMS	LPDAA C_ECS	NOAA_NCEI	NSIDC_ECS	NSID CV0	OB_D AAC	OMI NRT	ORNL_D AAC	POD AAC	SE DAC	USGS_EROS	USGS_LTA
Metadata Contact			.11																						
Distribution Contact	1.00		.89							1.00		.36	1.00		.96		.91		1.00		1.00	1.09	.92	1.00	
Point of Contact		2.97	.21	1.52	1.39	3.25	2.00	1.64	2.24		2.10	2.35	.01	2.00	.07	2.58		4.21					.23		1.93
Resource Author / Originator												.16			.82		1.49								
Processor																									

Feb 2016 Version