

UMM-Collection Metadata Completeness in the CMR

The UMM-Collection Profile consists of metadata concepts considered important for describing CMR data collections. The UMM-Collection Profile includes required, recommended and optional concepts. Required concepts must be included in UMM-compliant CMR metadata records.

The goal of this effort is to help NASA Data Centers provide full support for the UMM-Collection Profile through the evaluation and reporting of metadata completeness with respect to the UMM-Collection Profile. The following 26 CMR Metadata Collection (18 NASA and 8 IDN) were considered in this evaluation.

- Alaska Satellite Facility (ASF)
- Crustal Dynamics Data Information System (CDDIS)
- Global Hydrology Resource Center (GHRC)
- Goddard Earth Sciences Data and Information Center (GES_DISC)
- Level 1 and Atmosphere Archive and Distribution System (LAADS)
- Land, Atmosphere Near real-time Capability for EOS (LANCEMODIS)
- Land, Atmosphere Near real-time Capability for EOS (LANCEAMSR2)
- Langley Research Center (LARC)
- Langley Research Center (LARC_ASDC) Atmospheric Science Data Center
- Land Process DAAC - EOS Core System (LPDAAC_ECS)
- National Snow and Ice Data Center Version 0 (NSIDCV0)
- National Snow and Ice Data Center EOS Core System (NSIDC_ECS)
- Ocean Biology Processing Group (OBPG)
- Oak Ridge National Laboratory (ORNL)
- Ozone Monitoring Instrument Near Real Time (OMINRT)
- Physical Oceanography DAAC (PODAAC)
- Socioeconomic Data and Applications Center (SEDAC)
- U.S. Geological Survey Earth Resources Observation Systems (USGS_EROS)
- Australian Antarctic Data Centre (AU_AADC)
- European Space Agency (ESA)
- European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)
- Indian Space Research Organisation (ISRO)
- Japan Aerospace Exploration Agency (JAXA)
- LM_FIRMS
- NOAA's National Centers for Environmental Information (NCEI)
- U.S. Geological Survey Long Term Archive (USGS_LTA)

UMM-Collection Required Concepts in NASA Metadata Collections

The UMM-Collection profile identifies 15 concepts which are required for inclusion in CMR metadata collections. Table 1 shows the overall percent completeness of required UMM-Collection concepts in the 18 NASA metadata collections considered in this evaluation.

Table 1 - Percent Completeness in NASA Collections

Required Concept	% Complete	Required Concept	% Complete	Required Concept	% Complete	Required Concept	% Complete
Metadata Dates	100%	Abstract	100%	Keyword	100%	Platform Short Name	97%
Resource Identifier	100%	Data Dates	100%	Related URL	94%	Instrument Short Name	93%
Resource Version	100%	Responsibility	100%	Spatial Extent	95%	Project Name	73%
Resource Title	100%	Processing Level	99%	Temporal Extent	100%		

Specific Metadata Improvement Guidance

Of the 15 required concepts, 9 concepts are 100% complete in all CMR NASA collections, 5 concepts are > 90% complete in all CMR NASA collections, and 1 concept is > then 70% complete in all CMR NASA collections. Table 3 below provides detailed metadata improvement guidance for the 6 required concepts < 100% complete. The concept link in the first column connects to the concept element in the ISO Explorer guidance pages. The chart in the last column shows the collections that are missing the concepts as well as record count. This chart includes a link to a Google Sheets filtered display that shows the records for each collection that are missing the concept. The filtered display column header is bold, and the first two columns in the table show which collection and records are missing the concept.

Table 2 - Missing Required Concept Guidance (NASA Collections)

Concept	% Complete	Completeness Check XPath	Summary /Guidance	Collections Missing Concept
---------	------------	--------------------------	-------------------	-----------------------------

Processing Level	99%	<ul style="list-style-type: none">• <code>*/gmd:contentInfo*/gmd:processingLevelCode/gmd:MD_Identifier/gmd:code/*</code>	<ul style="list-style-type: none">• The processing level concept is missing in 5 collections and in 41 out of 6367 NASA metadata records.• A Processing Level value = 'Not provided' is being used in a varying capacity in 6 NASA collections (LANCEMODIS, LAADS, LPDAAC_ECS, LARC, NSIDC_ECS and SEDAC). In these cases we recommend replacing the 'Not Provided' value with <code>@nilReason='unknown'</code>	<table><tr><th>Collection</th><th>Count</th></tr><tr><td>USGS_EROS</td><td>5</td></tr><tr><td>OMINRT</td><td>5</td></tr><tr><td>NSIDCVO</td><td>1</td></tr><tr><td>NSIDC_ECS</td><td>28</td></tr><tr><td>LARC</td><td>1</td></tr></table>	Collection	Count	USGS_EROS	5	OMINRT	5	NSIDCVO	1	NSIDC_ECS	28	LARC	1				
Collection	Count																			
USGS_EROS	5																			
OMINRT	5																			
NSIDCVO	1																			
NSIDC_ECS	28																			
LARC	1																			
Platform Short Name	97%	<ul style="list-style-type: none">• <code>//gmi:platform*/gmi:identifier/gmd:MD_Identifier/gmd:code/*</code>	<ul style="list-style-type: none">• Platform Short Name is missing in 4 collections and in 207 out of 6367 NASA metadata records.• Platform Short Name is most commonly missing from SEDAC records. These records may not have a need for Platform documentation	<table><tr><th>Collection</th><th>Count</th></tr><tr><td>CDDIS</td><td>1</td></tr><tr><td>GHRC</td><td>1</td></tr><tr><td>ORNL_DAAC</td><td>10</td></tr><tr><td>SEDAC</td><td>180</td></tr></table>	Collection	Count	CDDIS	1	GHRC	1	ORNL_DAAC	10	SEDAC	180						
Collection	Count																			
CDDIS	1																			
GHRC	1																			
ORNL_DAAC	10																			
SEDAC	180																			
Spatial Extent	95%	<ul style="list-style-type: none">• <code>*/gmd:identificationInfo*/gmd:extent/gmd:EX_Extent/gmd:geographicElement/*</code>	<ul style="list-style-type: none">• Spatial Extent is missing in 7 collections and in 333 out of 6367 NASA metadata records.• Spatial Extent is most commonly missing from the LARC_ASDC records. It appears that 333 out of 606 LARC_ASDC records do not include a geographicElement.	<table><tr><th>Collection</th><th>Count</th></tr><tr><td>SEDAC</td><td>1</td></tr><tr><td>OMINRT</td><td>1</td></tr><tr><td>NSIDCVO</td><td>20</td></tr><tr><td>LARC_ASDC</td><td>280</td></tr><tr><td>LANCEAM...</td><td>1</td></tr><tr><td>GES_DISC</td><td>10</td></tr><tr><td>CDDIS</td><td>1</td></tr></table>	Collection	Count	SEDAC	1	OMINRT	1	NSIDCVO	20	LARC_ASDC	280	LANCEAM...	1	GES_DISC	10	CDDIS	1
Collection	Count																			
SEDAC	1																			
OMINRT	1																			
NSIDCVO	20																			
LARC_ASDC	280																			
LANCEAM...	1																			
GES_DISC	10																			
CDDIS	1																			
Related URL	94%	<ul style="list-style-type: none">• <code>//gmd:MD_DigitalTransferOptions/gmd:online/gmd:CI_OnlineResource/gmd:linkage/gmd:URL</code>• <code>*/gmd:identificationInfo*/gmd:graphicOverview/gmd:MD_BrowseGraphic/gmd:fileName/*</code>• <code>*/gmd:identificationInfo*/gmd:graphicOverview/gmd:MD_BrowseGraphic/gmx:fileName/*</code>	<ul style="list-style-type: none">• Related URL is missing in 5 collections and in 408 out of 6367 NASA metadata records.• Related URL is most commonly missing from the LARC, and LARC_ASDC metadata records. It appears that 91 of 407 LARC records do not include MD_DigitalTransferOptions content and that 297 of 696 LARC_ASDC records do not include MD_DigitalTransferOptions content.	<table><tr><th>Collection</th><th>Count</th></tr><tr><td>USGS_EROS</td><td>1</td></tr><tr><td>OMINRT</td><td>1</td></tr><tr><td>NSIDC_ECS</td><td>1</td></tr><tr><td>LARC_ASDC</td><td>280</td></tr><tr><td>LARC</td><td>91</td></tr></table>	Collection	Count	USGS_EROS	1	OMINRT	1	NSIDC_ECS	1	LARC_ASDC	280	LARC	91				
Collection	Count																			
USGS_EROS	1																			
OMINRT	1																			
NSIDC_ECS	1																			
LARC_ASDC	280																			
LARC	91																			

Instrument Short Name	93%	<ul style="list-style-type: none"><code>//gmi:instrument/*//gmi:identifier/gmd:MD_Identifier/gmd:code/*</code>	<ul style="list-style-type: none">Instrument Short Name is missing in 5 collections and in 420 out of 6367 NASA metadata records.Instrument Short Name is most commonly missing from the NSIDCV0 and SEDAC metadata records. It appears that 195 of 784 NSIDCV0 records do not include EOS_Instrument or MI_Instrument content and the majority of SEDAC records do not include EOS_Instrument or MI_Instrument content.	<table><thead><tr><th>Collection</th><th>Count</th></tr></thead><tbody><tr><td>CDDIS</td><td>5</td></tr><tr><td>GHRC</td><td>10</td></tr><tr><td>ORNL_DAAC</td><td>30</td></tr><tr><td>SEDAC</td><td>195</td></tr><tr><td>NSIDCV0</td><td>195</td></tr></tbody></table>	Collection	Count	CDDIS	5	GHRC	10	ORNL_DAAC	30	SEDAC	195	NSIDCV0	195												
Collection	Count																											
CDDIS	5																											
GHRC	10																											
ORNL_DAAC	30																											
SEDAC	195																											
NSIDCV0	195																											
Project Name	73%	<ul style="list-style-type: none"><code>/*//gmd:identificationInfo/*//gmd:aggregationInfo/gmd:MD_AggregateInformation[normalize-space(gmd:associationType/gmd:DS_AssociationTypeCode)='largerWorkCitation' and normalize-space(gmd:initiativeType/gmd:DS_InitiativeTypeCode)='project']/gmd:aggregateDataSetName/gmd:CI_Citation/gmd:title/*</code><code>/*//gmd:identificationInfo/*//gmd:descriptiveKeywords/gmd:MD_Keywords[normalize-space(gmd:type/gmd:MD_KeywordTypeCode)='project']/gmd:keyword/*</code><code>/*//gmi:acquisitionInformation/gmi:MI_AcquisitionInformation/gmi:operation/gmi:MI_Operation/gmi:citation/gmd:CI_Citation/*</code>	<ul style="list-style-type: none">Project Name is missing in 11 collections and in 1689 out of 6367 NASA metadata records.Project Name is the most commonly missing required UMM-Collection concept.The Project Name concept is checked for existence in 3 different metadata sections (Aggregation Info, Place Keyword and Operation).Project Name is most commonly found in the MI_Operation object in NASA metadata collections. The UMM-Recommendation view shows the collections that include and are missing this concept as well the concept occurrence count per collection.	<table><thead><tr><th>Collection</th><th>Count</th></tr></thead><tbody><tr><td>CDDIS</td><td>10</td></tr><tr><td>LAADS</td><td>10</td></tr><tr><td>OMINRT</td><td>10</td></tr><tr><td>USGS_EROS</td><td>10</td></tr><tr><td>LARC</td><td>80</td></tr><tr><td>ASF</td><td>100</td></tr><tr><td>OB_DAAC</td><td>120</td></tr><tr><td>NSIDC_ECS</td><td>150</td></tr><tr><td>LPDAAC_ECS</td><td>280</td></tr><tr><td>NSIDCV0</td><td>580</td></tr><tr><td>PODAAC</td><td>580</td></tr></tbody></table>	Collection	Count	CDDIS	10	LAADS	10	OMINRT	10	USGS_EROS	10	LARC	80	ASF	100	OB_DAAC	120	NSIDC_ECS	150	LPDAAC_ECS	280	NSIDCV0	580	PODAAC	580
Collection	Count																											
CDDIS	10																											
LAADS	10																											
OMINRT	10																											
USGS_EROS	10																											
LARC	80																											
ASF	100																											
OB_DAAC	120																											
NSIDC_ECS	150																											
LPDAAC_ECS	280																											
NSIDCV0	580																											
PODAAC	580																											

UMM-Collection Required Concepts in IDN Metadata Collections

The UMM-Collection profile identifies 15 concepts that are required for inclusion in CMR metadata collections. Table 2 shows the overall percent completeness of required UMM-Collection concepts in the 8 IDN metadata collections considered in this evaluation.

Table 3 - Percent Completeness in IDN Collections

Required Concept	% Complete
Metadata Dates	100%
Resource Identifier	100%
Resource Version	100%
Resource Title	100%

Abstract	100%
Data Dates	100%
Responsibility	100%
Processing Level	100%
Keyword	100%
Related URL	100%
Spatial Extent	99%
Temporal Extent	100%
Platform Short Name	100%
Instrument Short Name	60%
Project Name	26%

Specific Metadata Improvement Guidance

Of the 15 required concepts, 12 concepts are 100% complete in all CMR IDN collections, 1 concepts is > 90% complete in all CMR IDN collections, 1 concept is 60% complete in all CMR NASA collections, 1 concept is > 20% complete in all CMR IDN collections. Table 4 below provides detailed metadata improvement guidance for the 3 required concepts < 100% complete. The concept link in the first column connects to the concept element in the ISO Explorer guidance pages. The chart in the last column shows the collections that are missing the concepts as well as record count. This chart includes a link to a Google Sheets filtered display that shows the records for each collection that are missing the concept. The filtered display column header is bold, and the first two columns in the table show which collection and records are missing the concept.

Table 4 - Missing Required Concept Guidance (IDN Collections)

Concept	% Complete	Path	Guidance															
Spatial Extent	99%	<ul style="list-style-type: none">*/gmd:identificationInfo/* /gmd:extent/gmd:EX_Extent/gmd:geographicElement/*	<ul style="list-style-type: none">Spatial Extent is missing in 4 IDN collections and in 75 out of 8,702 IDN metadata records.Spatial Extent is missing in some capacity from JAXA, AU_AADC, NOAA_NCEI and ESA records.Each of these collection appear to be missing in small quantities geographicElement content.	<table><thead><tr><th>Collection</th><th>Records Missing</th></tr></thead><tbody><tr><td>AU_AADC</td><td>2</td></tr><tr><td>JAXA</td><td>5</td></tr><tr><td>ESA</td><td>10</td></tr><tr><td>NOAA_NCEI</td><td>55</td></tr></tbody></table>	Collection	Records Missing	AU_AADC	2	JAXA	5	ESA	10	NOAA_NCEI	55				
Collection	Records Missing																	
AU_AADC	2																	
JAXA	5																	
ESA	10																	
NOAA_NCEI	55																	
Instrument Short Name	60%	<ul style="list-style-type: none">//gmi:instrument/*/gmi:identifier/gmd:MD_Identifier/gmd:code/*	<ul style="list-style-type: none">Instrument Short Name is missing in 6 IDN collections and in 3,490 out of 8,702 IDN metadata records.Instrument Short Name is most commonly missing from AU_AADC and NOAA_NCEI collections. It appears that 1,720 of 2,559 AU_DAAC records and 1,719 of 5,488 NOAA_NCEI records are missing this concept	<table><thead><tr><th>Collection</th><th>Records Missing</th></tr></thead><tbody><tr><td>ESA</td><td>20</td></tr><tr><td>JAXA</td><td>50</td></tr><tr><td>EUMETSAT</td><td>100</td></tr><tr><td>USGS_LTA</td><td>1500</td></tr><tr><td>NOAA_NCEI</td><td>1700</td></tr><tr><td>AU_AADC</td><td>1700</td></tr></tbody></table>	Collection	Records Missing	ESA	20	JAXA	50	EUMETSAT	100	USGS_LTA	1500	NOAA_NCEI	1700	AU_AADC	1700
Collection	Records Missing																	
ESA	20																	
JAXA	50																	
EUMETSAT	100																	
USGS_LTA	1500																	
NOAA_NCEI	1700																	
AU_AADC	1700																	

Project Name	26%	<ul style="list-style-type: none">• <code>/*gmd:identificationInfo/* /gmd:aggregationInfo/gmd: MD_AggregateInformation [normalize-space(gmd: associationType/gmd: DS_AssociationTypeCode) ='largerWorkCitation' and normalize-space(gmd: initiativeType/gmd: DS_InitiativeTypeCode) ='project']/gmd: aggregateDataSetName /gmd:CI_Citation/gmd:title /*</code>• <code>/*gmd:identificationInfo/* /gmd:descriptiveKeywords /gmd:MD_Keywords [normalize-space(gmd: type/gmd: MD_KeywordTypeCode) ='project']/gmd:keyword/*</code>• <code>/*gmi: acquisitionInformation/gmi: MI_AcquisitionInformation /gmi:operation/gmi: MI_Operation/gmi:citation /gmd:CI_Citation/*</code>	<ul style="list-style-type: none">• Project Name is missing in 6 IDN collections and in 6,419 out of 8,702 IDN metadata records.• Project Name is the most commonly missing required UMM-Collection concept.• The Project Name concept is checked for existence in 3 different metadata sections (Aggregation Info, Place Keyword and Operation).• It appears that the Project Name is most commonly found in the MI_Operation object in IDN metadata collections. The UMM-Recommendation view shows the collections that include and are missing this concept, as well the concept occurrence count per collection.• Project Name is 100% complete in the ISRO metadata collection with documentation occurring in the MD_Operation object.	<table><tr><th>Collection</th><th>Count</th></tr><tr><td>EUMETSAT</td><td>~100</td></tr><tr><td>USGS_LTA</td><td>~100</td></tr><tr><td>ESA</td><td>~100</td></tr><tr><td>JAXA</td><td>~400</td></tr><tr><td>AU_AADC</td><td>~2400</td></tr><tr><td>NOAA_NCEI</td><td>~3500</td></tr></table>	Collection	Count	EUMETSAT	~100	USGS_LTA	~100	ESA	~100	JAXA	~400	AU_AADC	~2400	NOAA_NCEI	~3500
Collection	Count																	
EUMETSAT	~100																	
USGS_LTA	~100																	
ESA	~100																	
JAXA	~400																	
AU_AADC	~2400																	
NOAA_NCEI	~3500																	