

Collection Progress

- [Element Description](#)
- [Best Practices](#)
- [Element Specification](#)
- [Metadata Validation and QA/QC](#)
- [Dialect Mappings](#)
 - [DIF 10](#)
 - [ECHO 10](#)
 - [ISO 19115-2 MENDS](#)
 - [ISO 19115-2 SMAP](#)
- [UMM Migration](#)
- [History](#)
 - [UMM Versioning](#)
 - [ARC Documentation](#)

Element Description

The Collection Progress element describes the production status of the dataset. The Collection Progress element leverages a controlled vocabulary to ensure consistency across CMR. There are five possible choices for describing the status of the dataset:

- **PLANNED** refers to datasets to be collected in the future and are thus unavailable at the present time. Examples include:
 - The Hydro spacecraft has not been launched, but information on planned datasets may be available.
- **ACTIVE** refers to datasets currently in production or data that is continuously being collected or updated. Examples include:
 - Data from an instrument that continually makes observations such as the AIRS instrument on Aqua or MODIS on Terra.
 - Datasets where one version of a dataset is continuously and regularly updated such as VIIRS/NPP Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V001 (doi: 10.5067/VIIRS/VNP09GA.001)
- **COMPLETE** refers to datasets in which no updates or further data collection will be made. Examples include:
 - Data collection from the Lightning Imaging Sensor (LIS) has been completed due to the end of the TRMM mission.
 - Completion of a legacy version of a product where no further updates will be made such as with CERES EBAF-TOA Ed2.7.
- **DEPRECATED**: Deprecated products have been retired but are still discoverable for historical purposes.
- **NOT APPLICABLE** should only be used if this element is not applicable to the collection, such as a calibration collection.

Best Practices

For continuous datasets:

If data collection is ongoing and the Collection Progress element is set to "ACTIVE", the following actions are recommended:

- The 'Ends at Present Flag' element should be set to "true."
- If the [temporal extent](#) of the collection is expressed as a range date time, then it is not necessary to populate the "Ending Date Time" element in the metadata.
 - For additional information see the [Temporal Extent](#) wiki page

Setting the 'Ends at Present Flag' element to "true" tells the CMR that the ending time for the collection is present day, and thus eliminates the need to specify the ending date time of the collection. This also eliminates the need to update the ending date time in the metadata each time new data gets added to the collection.

For completed datasets:

If data collection is complete and the collection progress is set to "COMPLETE", the following actions are recommended:

- The 'Ends at Present Flag' element should be set to "false". Alternatively, the 'Ends at Present Flag' element may be completely removed from the metadata since it is an optional element.
- If the [temporal extent](#) of the collection is expressed as a range date time, then the "Ending Date Time" element must be provided in the metadata.
 - For additional information see the [Temporal Extent](#) wiki page

Setting the 'Ends at Present Flag' element to "false" tells the CMR that the ending time for the collection is in the past. If the temporal extent of the collection is expressed as a range date time, then the "Ending Date Time" element should specify the ending date and time of the last available granule in the collection.

For disparate datasets:

For some datasets, there may be gaps in data collection. For example, there may be a flight campaign dataset where data is only collected in May and September of each year. If there are future plans to add data to the collection, then it is okay to set the Collection Progress to "ACTIVE". In this scenario, opposite to the best practices specified for continuous datasets above, it is recommended that the 'Ends At Present Flag' be set to "false" and that an Ending Date Time be provided. This would require that the Ending Date Time be updated each time new data gets added to the collection (e.g. in May and September). This practice most accurately conveys the temporal coverage of a dataset to a user.

Element Specification

Collection Progress is required. Only one Collection Progress value may be provided (Cardinality: 1)

Model	Element	Type	Constraints	Required?	Cardinality
UMM-C	CollectionProgress	Enumeration	PLANNED ACTIVE COMPLETE DEPRECATED NOT APPLICABLE	Yes	1

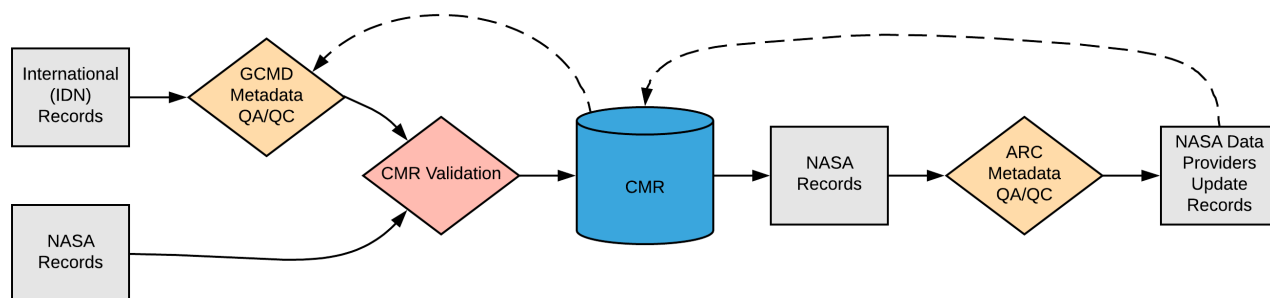
Value needed for translations:

The following value is needed by the CMR to translate older non UMM compliant records to and from the UMM and other supported specifications where no valid value is given in a field required by the UMM. This is needed partly because the CMR still allows a non UMM compliant record to be ingested with warnings.

NOT PROVIDED - It is necessary for this value to exist so that the CMR can translate older non UMM compliant records into the latest UMM specification where CollectionProgress is required. *This value should not be used by metadata providers.*

Metadata Validation and QA/QC

All metadata entering the CMR goes through the below process to ensure metadata quality requirements are met. All records undergo CMR validation before entering the system. The process of QA/QC is slightly different for NASA and non-NASA data providers. Non-NASA providers include interagency and international data providers and are referred to as the International Directory Network (IDN).



Please see the expandable sections below for flowchart details.

- Manual Review
 - Identify errors, discrepancies or omissions.
- Automated Review
 - Check that the field has been populated.
 - Check that the field value is valid.

Automated Review

- Check that the field has been populated.
- Check that the field value is valid.

ARC Priority Matrix

Priority Categorization	Justification
-------------------------	---------------

Red = High Priority Finding	This element is categorized as highest priority when: - No Collection Progress is provided. - An invalid value is provided for Collection Progress. Valid values include: PLANNED, ACTIVE, COMPLETE, DEPRECATED, NOT APPLICABLE - The Collection Progress appears to be out of sync with data collection. For example: - Data collection stopped in the distant past but progress is listed as 'ACTIVE' - as a rule of thumb, this applies when the last available granule has an ending date time of 1+ years in the past. - Collection Progress is listed as 'PLANNED' but data is actively being collected. - Data collection is ongoing but the element lists the progress as 'COMPLETE'
Yellow = Medium Priority Finding	This element is categorized as medium priority when: - The Collection Progress appears to be out of sync with data collection. For example: - The ending date time of the latest granule in the collection is in the past, however the Collection Progress is 'ACTIVE' - this will be marked yellow when the latest granule in the collection is less than 1 year from the present day, and the collection is part of a field or flight campaign which may still be ongoing (this could result in gaps in the data). The DAAC should confirm whether data collection is still ongoing for the field/flight campaign or whether the collection is complete.
Blue = Low Priority Finding	Not Applicable
Green = No Findings/Issues	The element is provided and follows all applicable criteria specified in the best practices section above.

ARC Automated Checks

ARC uses the [pyQuARC library](#) for automated metadata checks. Please see the [pyQuARC GitHub](#) for more information.

Dialect Mappings

DIF 9 (Note: DIF-9 is being phased out and will no longer be supported after 2018)

DIF 10

Dataset_Progress is optional in DIF 10. Only one **Dataset_Progress** value may be provided (Cardinality: 0..1)

UMM-C Element	DIF 10 Path	Type	Constraints	Required in DIF 10?	Cardinality	Notes
CollectionProgress	/DIF /Dataset_Progress	Enumeration	PLANNED IN WORK COMPLETE	No	0..1	Dataset_Progress is optional in DIF 10, however, it is required for EOSDIS datasets.

Enumeration Mapping

DIF 10	Translation Direction	UMM
PLANNED		PLANNED
IN WORK		ACTIVE
COMPLETE		COMPLETE
COMPLETE		DEPRECATED
Blank or doesn't exist		NOT PROVIDED
Don't translate		NOT PROVIDED

Example Mapping

DIF 10

```
<Dataset_Progress>COMPLETE</Dataset_Progress>
```

UMM

```
"CollectionProgress" : "COMPLETE",
```

ECHO 10

Collection State is optional in ECHO 10. Only one Collection State value may be provided (Cardinality: 0..1)

UMM-C Element	ECHO 10 Path	Type	Constraints	Required in ECHO10?	Cardinality	Notes
CollectionProgress	Collection State	String	1 - 80 characters	No	0..1	While this field is not enumeration controlled in ECHO 10, use of the UMM enumeration values (PLANNED, ACTIVE, COMPLETE, DEPRECATED, NOT APPLICABLE) is strongly encouraged in order to prevent translation errors. This field is required for EOSDIS datasets.

Enumeration Mapping

ECHO 10	Translation Direction	UMM
PLANNED		PLANNED
IN WORK		ACTIVE
COMPLETE		COMPLETE
completed		COMPLETE
COMPLETE		DEPRECATED
NOT APPLICABLE		NOT APPLICABLE
Blank or doesn't exist		NOT PROVIDED
Any other value		NOT PROVIDED
Don't translate		NOT PROVIDED

Example Mapping

ECHO 10

```
<CollectionState>COMPLETED</CollectionState>
```

UMM

```
"CollectionProgress" : "COMPLETE",
```

ISO 19115-2 MENDS

This field is optional in ISO. Multiple Collection Progress values may be provided (Cardinality: 0..*). Note: only the first value provided will be translated to the CMR.

UMM-C Element	ISO Path	Type	Notes
CollectionProgress	/gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:status/gmd:MD_ProgressCode codeList="https://cdn.earthdata.nasa.gov/iso/resources/Codelist/gmx/Codelists.xml#MD_ProgressCode" codeListValue=	String	The ProgressCode has code values of: completed, historicalArchive, obsolete, onGoing, planned, required, underDevelopment, final, pending, retired, superseded, tentative, valid, accepted, notAccepted, withdrawn, proposed, and deprecated. This field is optional in ISO, but is required to be provided for any EOSDIS dataset. Any string can be substituted as well. Since ISO supports multiple statuses for a collection/series, the CMR translates only the first one to UMM.

Enumeration/Code List Mapping

ISO MENDS	Translation Direction	UMM
planned		PLANNED
underDevelopment		PLANNED
onGoing		ACTIVE
completed		COMPLETE
historicalArchive		COMPLETE
obsolete		COMPLETE
retired		DEPRECATED
deprecated		DEPRECATED
NOT APPLICABLE a string is used instead of the defined codes. The codeList="" and codeListValue = ""		NOT APPLICABLE
Blank or doesn't exist		NOT PROVIDED
Any other value		NOT PROVIDED
Don't translate		NOT PROVIDED

Example Mapping

ISO 19115-2 MENDS

```
<gmd:status>
  <gmd:MD_ProgressCode codeList=
    "https://cdn.earthdata.nasa.gov/iso/resources
    /CodeList/gmxCodeLists.xml#MD_ProgressCode"
    codeListValue="completed">completed</gmd:
MD_ProgressCode>
</gmd:status>
```

UMM

```
"CollectionProgress" : "COMPLETE",
```

ISO 19115-2 SMAP

This field is optional in ISO. Multiple Collection Progress values may be provided (Cardinality: 0..*). Note: only the first value provided will be translated to the CMR.

UMM-C Element	ISO Path	Type	Notes
---------------	----------	------	-------

CollectionProgress	/gmd:DS_Series/gmd:seriesMetadata/gmi:MI_Metadata/gmd:identificationInfo/ gmd:MD_DataIdentification /gmd:status/gmd:MD_ProgressCode codeList="https://cdn.earthdata.nasa.gov/iso/resources/Codelist/gmxCodeLists.xml#MD_ProgressCode" codeListValue=	String	The ProgressCode has code values of: completed, historicalArchive, obsolete, onGoing, planned, required, underDevelopment, final, pending, retired, superseded, tentative, valid, accepted, notAccepted, withdrawn, proposed, and deprecated. This field is optional in ISO, but is required to be provided for any EOSDIS dataset. Any string can be substituted as well. Since ISO supports multiple statuses for a collection/series, the CMR translates only the first one to UMM.
--------------------	--	--------	--

Enumeration/Code List Mapping

ISO SMAP	Translation Direction	UMM
planned		PLANNED
underDevelopment		PLANNED
onGoing		ACTIVE
completed		COMPLETE
historicalArchive		COMPLETE
obsolete		COMPLETE
retired		DEPRECATED
deprecated		DEPRECATED
NOT APPLICABLE a string is used instead of the defined codes. The codeList="" and codeListValue = ""		NOT APPLICABLE
Blank or doesn't exist		NOT PROVIDED
Any other value		NOT PROVIDED
Don't translate		NOT PROVIDED

Example Mapping

ISO 19115-2 SMAP

```
<gmd:status>
  <gmd:MD_ProgressCode codeList=
    "https://cdn.earthdata.nasa.gov/iso
    /resources/Codelist/gmxCodeLists.xml#MD_ProgressCode"
    codeListValue="completed">completed</gmd:
MD_ProgressCode>
</gmd:status>
```

UMM

```
"CollectionProgress" : "COMPLETE",
```

UMM Migration

UMM Version 1.9.0	Translation Direction	UMM Version 1.10.0
PLANNED		PLANNED
IN WORK		ACTIVE
COMPLETE		COMPLETE
COMPLETE		DEPRECATED
NOT APPLICABLE		NOT APPLICABLE
NOT PROVIDED		NOT PROVIDED
Any other value		NOT PROVIDED

History

UMM Versioning

Version	Date	What Changed
1.15.5	12/3/2020	No changes were made for Collection Progress during the transition from version 1.15.4 to 1.15.5
1.15.4	9/18/2020	No changes were made for Collection Progress during the transition from version 1.15.3 to 1.15.4
1.15.3	7/1/2020	No changes were made for Collection Progress during the transition from version 1.15.2 to 1.15.3
1.15.2	5/20/2020	No changes were made for Collection Progress during the transition from version 1.15.1 to 1.15.2
1.15.1	3/25/2020	Added the DEPRECATED enumeration to Collection Progress.
1.15.0	2/26/2020	No changes were made for Collection Progress during the transition from version 1.14.0 to 1.15.0
1.14.0	10/21/2019	No changes were made for Collection Progress during the transition from version 1.13.0 to 1.14.0
1.13.0	04/11/2019	No changes were made for Collection Progress during the transition from version 1.12.0 to 1.13.0
1.12.0	01/22/2019	No changes were made for Collection Progress during the transition from version 1.11.0 to 1.12.0.
1.11.0	11/28/2018	No changes were made for Collection Progress during the transition from version 1.10.0 to 1.11.0.
1.10.0	05/02/2018	During the transition from version 1.9.0 to 1.10.0, the enumeration value "IN WORK" was changed to "ACTIVE."
1.9.0	12/04/2017	n/a

ARC Documentation

Version	Date	What Changed	Author
1.0	02/19/18	Recommendations/priority matrix transferred from internal ARC documentation to wiki space	Kaylin Bugbee