

Archive and Distribution Information for Collections

- [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

Archive and distribution information elements provide information about a format other than its original for data immediately available for download. This includes information such as the average downloadable file size, data format, mime type, checksum, and an estimate of the collection's entire downloadable file size.

Best Practices

There are multiple sub-elements that comprise Archive and Distribution Information for Collections:

Format: Defines a single format for an archival artifact (e.g., binary, jpeg, HDF5, HDF-EOS5, GeoTIFF). The format must be selected from the [GCMD Granule Data Format vocabulary list](#) in order to adhere to the UMM-C schema. This ensures consistent use of data formats across the CMR.

Format Type: Allows the provider to define if the archival artifact's format is in its native format or another supported format.

Format Description: Allows the data provider to provide added information about the provided Format.

Average File Size: An approximate average of the size of the the archivable item, which gives users an idea of the magnitude of each archivable file.

Average File Size Unit: Unit of measure for the average file size (e.g., KB, MB, TB).

Total Collection File Size: An approximate total size of the archivable items, which gives users an idea of the magnitude of all the archivable files combined.

Total Collection File Size Unit: Unit of measure for the total collection file size (e.g., KB, MB, TB).

Total Collection File Size Begin Date: The date that data collection began for the collection, which should be in the following format: yyyy-MM-ddTHH:mm:ssZ

Description: Enables the provider to provide more information about the archivable item.

Media: Defines how the distributable material can be obtained by an end user (e.g., Online Archive, CD-ROM, hard drives, etc.).

Fees: Provide the price an end user will need to pay in order to obtain the distributable material.

Each of these elements are available under 'FileArchiveInformation' and 'FileDistributionInformation'. **FileArchiveInformation** describes the data in its archived format. **FileDistributionInformation** describes the data as distributed by a data provider. It can be the original format in which the data are archived, or it can be used to specify when the data is distributed in a format other than its original format. **FileDistributionInformation is the preferred location for providing the data format since this should specify the format data are being distributed in to the end user.**

Examples:

Format: "HDF-EOS5"

Format Type: "Native"

Format Description: "This is an EOSDIS specific format, and it is also available in ASCII."

Average File Size: "250"

Average File Size Unit: "KB"

Total Collection File Size: "25"

Total Collect File Size Unit: "GB"

Total Collection File Size Begin Date: "2018-04-01T00:00:00Z"

Element Specification

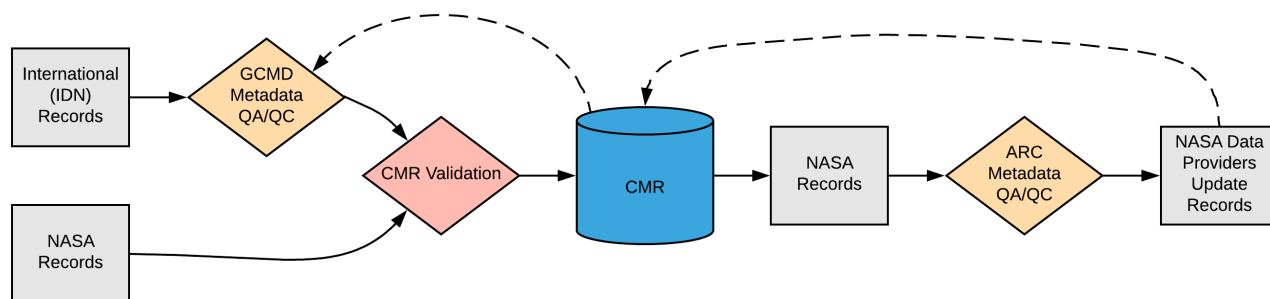
Cardinality of Archive and Distribution Information:

Model	Element	Type	Usable Valid Values	Constraints	Required?	Cardinality	Notes
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/ Format	String	GCMD Granule Data Format vocabulary list	1 - 80 characters	Yes	1	The format must be selected from the GCMD Granule Data Format vocabulary list .
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/FormatType	String	Native Supported	n/a	No	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/FormatDescription	String		1 - 80 characters	No	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/ AverageFileSize	Decimal			Yes, if TotalCollectionFileSizeBeginDate is used.	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/ AverageFileSize Unit	String	KB MB GB TB PB NA	n/a	Yes, if AverageFileSize is used.	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/ TotalCollectionFileSize	Decimal			No	0..1	Either use TotalCollectionFileSize or TotalCollectionFileSizeBeginDate
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/ TotalCollectionFileSizeUnit	Enumeration	KB MB GB TB PB NA	n/a	Yes, if TotalCollectionFileSize is used	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/ TotalCollectionFileSizeBeginDate	Date			No	0..1	Either use TotalCollectionFileSize or TotalCollectionFileSizeBeginDate
UMM-C	ArchiveAndDistributionInformation/ FileArchiveInformation/ Description	String	n/a	1 - 1024 characters	No	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/Format	String	GCMD Granule Data Format vocabulary list	1 - 80 characters	Yes	1	The format must be selected from the GCMD Granule Data Format vocabulary list . Recommend providing the data format in this element since this should specify the format data are being distributed in to the end user.
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ FormatType	String	Native Supported		No	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ FormatDescription	String		1 - 80 characters	No	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ Media	String	n/a	1 - 80 characters	No	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ AverageFileSize	Decimal	n/a	n/a	Yes, if TotalCollectionFileSizeBeginDate is used.	0..1	

UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ AverageFileSizeUnit	Enumeration	KB MB GB TB PB NA		Yes, if AverageFileSize is used.	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ TotalCollectionFileSize	Decimal	n/a		No	0..1	Either use TotalCollectionFileSize or TotalCollectionFileSizeBeginDate
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ TotalCollectionFileSizeUnit	Enumeration	KB MB GB TB PB NA	n/a	Yes, if TotalCollectionFileSize is used	0..1	
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ TotalCollectionFileSizeBeginDate	Date			No	0..1	Either use TotalCollectionFileSize or TotalCollectionFileSizeBeginDate
UMM-C	ArchiveAndDistributionInformation/ FileDistributionInformation/ Description	String	n/a	1 - 1024 characters	No	0..1	
UMM-C	Fees	String	n/a	1 - 255 characters	No	0..1	

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: TBD
- Automated Review: TBD
- TBD

ARC Priority Matrix

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

Red = High Priority Finding	This element is categorized as highest priority when: - No Data Format is provided. - The Data Format provided is incorrect for the dataset. - The Data Format does not comply with the GCMD Granule Data Format vocabulary list. - The Data Format provided is correct, however, it does not exactly match its entry in the GCMD Granule Data Format vocabulary list. - E.g. providing 'netcdf4' versus 'netCDF-4' - E.g. providing 'geo-tiff' versus 'GeoTIFF'
Yellow = Medium Priority Finding	This element is categorized as medium priority when: - The Data Format provided is correct, but could be more specific. - E.g. Providing 'ASCII' when the data format should be 'CSV' - A recommendation is made to add a data format to the existing list.
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

All of the sub-elements are optional in DIF 10.

UMM-C Element	Path	Type	Usable Valid Values	Constraints	Required in DIF 10?	Cardinality	Notes
ArchiveAndDistributionInformation /FileDistributionInformation/ Format	/DIF/Distribution /Distribution_Format	String			No	0..1	To ensure consistency, it is highly recommended that the format be selected from the GCMD Granule Data Format vocabulary list .
ArchiveAndDistributionInformation /FileDistributionInformation/FormatType	/DIF/Distribution /Distribution_Format_Type	String	Native Supported		No	0..1	This is a DIF 10.3 field. Currently in DIF 10.2 there is just one field called ' /DIF/Distribution/Distribution_Format'
ArchiveAndDistributionInformation /FileDistributionInformation/Media	/DIF/Distribution /Distribution_Media	String			No	0..1	
ArchiveAndDistributionInformation /FileDistributionInformation /AverageFileSize	/DIF/Distribution /Average_Granule_Size						This is a DIF 10.3 field. Currently in DIF 10.2 there is just one field called ' /DIF/Distribution/Distribution_Size'
ArchiveAndDistributionInformation /FileDistributionInformation /AverageFileSizeUnit	/DIF/Distribution /Average_Granule_Size_Unit	Enumeration	KB MB GB TB PB		No	0..1	This is a DIF 10.3 field. Currently in DIF 10.2 there is just one field called ' /DIF/Distribution/Distribution_Size'
ArchiveAndDistributionInformation /FileDistributionInformation /TotalCollectionFileSize	/DIF/Distribution /Total_Collection_Size	Decimal			No	0..1	This is a DIF 10.3 field. Currently in DIF 10.2 there is just one field called ' /DIF/Distribution/Distribution_Size'
ArchiveAndDistributionInformation /FileDistributionInformation /TotalCollectionFileSizeUnit	/DIF/Distribution /Distribution_Collection_Size_Unit	Enumeration	KB MB GB TB PB		No	0..1	This is a DIF 10.3 field. Currently in DIF 10.2 there is just one field called ' /DIF/Distribution/Distribution_Size'
ArchiveAndDistributionInformation /FileDistributionInformation/ Fees	/DIF/Distribution/Fees	String	n/a		No	0..1	

Example Mapping

DIF 10

```

<Distribution>
  <Distribution_Media>Online Archive<
/Distribution_Media>
  <Average_Granule_Size>231</Average_Granule_Size>
  **Note: In DIF 10.2 this would appear as:
  <Distribution_Size>231 MB per file<
/Distribution_Size>
  <Average_Granule_Size_Unit>MB<
/Average_Granule_Size_Unit>
  <Distribution_Format>HDF-EOS5</Distribution_Format>
  <Distribution_Format_Type>Native<
/Distribution_Format_Type>  **Note: The
Distribution_Format_Type field is not available in
DIF 10.2
  <Fees>Free</Fees>
</Distribution>

```

UMM

```

"ArchiveAndDistributionInformation": {
  "FileDistributionInformation": [
    {
      "FormatType": "Native",
      "AverageFileSize": 231,
      "Fees": "Free",
      "Format": "HDF-EOS",
      "AverageFileSizeUnit": "MB",
      "Media": [
        "Online Archive"
      ]
    }
  ]
},

```

ECHO 10

Only 2 UMM-C Archive and Distribution Information sub-elements map to ECHO 10, Data Format and Fees, and neither element is required.

UMM-C Element	Path	Type	Usable Valid Values	Constraints	Required in ECHO10?	Cardinality	Notes
ArchiveAndDistributionInformation /FileDistributionInformation/Format	/DataFormat	String		N/A	No	0..1	To ensure consistency, it is highly recommended that the format be selected from the GCMD Granule Data Format vocabulary list .
ArchiveAndDistributionInformation /FileDistributionInformation/Fees	/Price	String		N/A	No	0..1	

Example Mapping

ECHO 10

```

<DataFormat>HDF-EOS5</DataFormat>
<Price>0.0</Price>

```

UMM

```

"ArchiveAndDistributionInformation": {
  "FileDistributionInformation": [
    {
      "FormatType": "Native",
      "Fees": "0.0",
      "Format": "HDF-EOS5"
    }
  ]
},

```

ISO 19115-2 MENDS

Archive and Distribution Information is an optional metadata element in ISO 19115-2 MENDS.

UMM-C Element	Path	Type	Notes
ArchiveAndDistributionInformation/FileArchiveInformation/Format	[=> /gmd:name/gco:CharacterString	String	
ArchiveAndDistributionInformation/FileArchiveInformation/FormatType	[=> /gmd:specification/gco:CharacterString = FormatType: <FormatType>	String	
ArchiveAndDistributionInformation/FileArchiveInformation/AverageFileSize	[=> /gmd:specification/gco:CharacterString = AverageFileSize: <AverageFileSize>	Decimal	
ArchiveAndDistributionInformation/FileArchiveInformation/AverageFileSizeUnit	[=> /gmd:specification/gco:CharacterString = AverageFileSizeUnit: <AverageFileSizeUnit>	String	
ArchiveAndDistributionInformation/FileArchiveInformation/TotalCollectionFileSize	[=> /gmd:specification/gco:CharacterString = TotalCollectionFileSize: <TotalCollectionFileSize>	Decimal	
ArchiveAndDistributionInformation/FileArchiveInformation/TotalCollectionFileSizeUnit	[=> /gmd:specification/gco:CharacterString = TotalCollectionFileSizeUnit: <TotalCollectionFileSizeUnit>	Enumeration	
ArchiveAndDistributionInformation/FileArchiveInformation/TotalCollectionFileSizeBeginDate	[=> /gmd:specification/gco:CharacterString = TotalCollectionFileSizeBeginDate: <TotalCollectionFileSizeBeginDate>	Date	
ArchiveAndDistributionInformation/FileArchiveInformation/Description	[=> /gmd:specification/gco:CharacterString = Description: <Description>	String	
ArchiveAndDistributionInformation/FileDistributionInformation/Format	[=> /gmd:distributionFormat xlink:href="FileDistributionInformation_<Block Number>/gmd:MD_Format/gmd:name/gco:CharacterString	String	
ArchiveAndDistributionInformation/FileDistributionInformation/FormatType	[=> /gmd:distributionFormat xlink:href="FileDistributionInformation_<Block Number>/gmd:MD_Format/gmd:specification/gmd:CharacterString = FormatType: <Native or Supported>	String	
ArchiveAndDistributionInformation/FileDistributionInformation/Media	[=> /gmd:transferOptions xlink:href="FileDistributionInformation_Media_<Block Number>/gmd:MD_DigitalTransferOptions/gmd:offLine/gmd:MD_Medium/gmd:name/gmd:MD_MediumNameCode codeList="https://cdn.earthdata.nasa.gov/iso/resources/Codelist/gmxCodetlists.xml#MD_MediumNameCode" codeListValue="<value>" = <value>	String	
ArchiveAndDistributionInformation/FileDistributionInformation/AverageFileSize	[=> /gmd:transferOptions xlink:href="FileDistributionInformation_AverageFileSize_<Block Number>/gmd:MD_DigitalTransferOptions/gmd:unitsOfDistribution/gco:CharacterString	Decimal	
ArchiveAndDistributionInformation/FileDistributionInformation/AverageFileSizeUnit	[=> /gmd:transferOptions xlink:href="FileDistributionInformation_AverageFileSize_<Block Number>/gmd:MD_DigitalTransferOptions/gmd:unitsOfDistribution/gco:CharacterString	Enumeration	
ArchiveAndDistributionInformation/FileDistributionInformation/TotalCollectionFileSize	[=> /gmd:transferOptions xlink:href="FileDistributionInformation_TotalCollectionFileSize_<Block Number>/gmd:MD_DigitalTransferOptions/gmd:transferSize/gco:Real	Decimal	
ArchiveAndDistributionInformation/FileDistributionInformation/TotalCollectionFileSizeUnit	[=> /gmd:transferOptions xlink:href="FileDistributionInformation_TotalCollectionFileSize_<Block Number>/gmd:MD_DigitalTransferOptions/gmd:unitsOfDistribution/gco:CharacterString	Enumeration	

ArchiveAndDistributionInformation/FileDistributionInformation/TotalCollectionFileSizeBeginDate	[=> /gmd:distributionFormat xlink:href="FileDistributionInformation_<Block Number>"/gmd:MD_Format/gmd:specification/gmd:CharacterString = TotalCollectionFileSizeBeginDate: <Date use MMDDYYYYTHH:MM:SSZ format>	Date	
ArchiveAndDistributionInformation/FileDistributionInformation/Description	[=> /gmd:distributor xlink:href="FileDistributionInformation_<Block Number>"/gmd:MD_Distributor/gmd:distributionOrderProcess/gmd:MD_StandardOrderProcess/gmd:orderingInstructions/gmd:CharacterString	String	
ArchiveAndDistributionInformation/FileDistributionInformation/Fees	[=> /gmd:distributor xlink:href="FileDistributionInformation_<Block Number>"/gmd:MD_Distributor/gmd:distributionOrderProcess/gmd:MD_StandardOrderProcess/gmd:fees/gmd:CharacterString	String	

Example Mapping

ISO 19115-2 MENDS

```
<gmi:MI_Metadata>
  ...
  <gmd:distributionFormat xlink:href="
FileDistributionInformation_<Block Number>"/>
    <gmd:MD_Format>
      <gmd:specification>
        <gco:
CharacterString>TotalCollectionFileSizeBeginDate<
/gco:CharacterString>
          </gmd:code>
          <gmd:codeSpace>
            <gco:CharacterString>2018-04-01T00:
00:00Z</gco:CharacterString>
          </gmd:codeSpace>
          <gmd:specification>
            <gmd:MD_Format>
              <gmd:distributionFormat xlink:href="
FileDistributionInformation_<Block Number>"/>
```

UMM

```
ArchiveandDistributionInformation: [
{
  AverageFileSize: "250",
},
{
  AverageFileSizeUnit: "KB",
},
{
  TotalCollectionFileSizeBeginDate: "2018-04-
01T00:00:00Z",
}
],
```

ISO 19115-2 SMAP

Archive and Distribution Information does not currently map to ISO 19115-2 SMAP.

UMM-C Element	Path	Type	Notes
N/A	N/A	N/A	N/A

Example Mapping

N/A
N/A

UMM Migration

None

History

UMM Versioning

Version	Date	What Changed
1.15.5	12/3/2020	The description of the format elements now includes text about the controlled vocabulary for this field coming from KMS.
1.15.4	9/18/2020	No changes were made for Archive and Distribution Information during the transition from version 1.15.3 to 1.15.4
1.15.3	7/1/2020	FormatDescription was added to FileArchive and FileDistribution Information types to version 1.15.3
1.15.2	5/20/2020	No changes were made for Archive and Distribution Information during the transition from version 1.15.1 to 1.15.2
1.15.1	3/25/2020	No changes were made for Archive and Distribution Information during the transition from version 1.15.0 to 1.15.1
1.15.0	2/26/2020	No changes were made for Archive and Distribution Information during the transition from version 1.14.0 to 1.15.0
1.14.0	10/21/2019	No changes were made for Archive and Distribution Information during the transition from version 1.13.0 to 1.14.0
1.13.0	04/11/2019	This element was added to the UMM-C specification.

ARC Documentation

Version	Date	What Changed	Author
1.0	05/16/18	Recommendations/priority matrix transferred from internal ARC documentation to wiki space	Jeanne'le Roux
1.1	04/12/21	Moved "The data format provided is correct, however, it does not exactly match its entry in the GCMD Granule Data Format vocabulary list ." from blue to red. GCMD keywords are now the authoritative controlled vocabulary source for data formats.	Jeanne'le Roux