

# Data Language

- [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 Data Language element describes the language used in the preparation, storage, and description of the collection. It is the language of the collection data itself. It does not refer to the language used in the metadata record (although this may be the same language). Please refer to the Metadata Language wiki page on how to specify the language used in the metadata.

## Best Practices

While not required, it is recommended that an ISO 639-2 code be used to populate the Data Language field ([http://www.loc.gov/standards/iso639-2/php/code\\_list.php](http://www.loc.gov/standards/iso639-2/php/code_list.php)). If a language is not supplied - English, the default language, will be assumed.

### Examples:

"eng" Example using the ISO 639-2 code for English

"French" Example where the ISO 639-2 code is not used

"chi (B)" Example using the ISO 639-2 code for Chinese (See [this FAQ](#) for an explanation why "(B)" is included in the code)

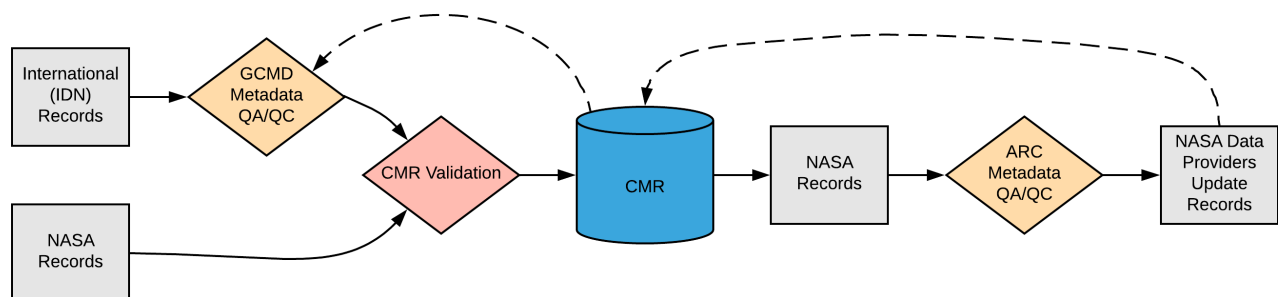
## Element Specification

Providing a Data Language is optional. Multiple Data Languages may be provided, if necessary (Cardinality: 0..\*)

Model	Element	Type	Constraints	Required?	Cardinality
UMM-C	DataLanguage	String	1 - 25 characters	No	0..*

## 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 value matches the enumeration value. (English;Afrikaans;Arabic;Bosnian;Bulgarian;Chinese;Croatian;Czech;Danish;Dutch;Estonian;Finnish;French;German;Hebrew;Hungarian;Indonesian;Italian;Japanese;Korean;Latvian;Lithuanian;Norwegian;Polish;Portuguese;Romanian;Russian;Slovak;Spanish;Ukrainian;Vietnamese)
- This element is currently not validated.

#### ARC Priority Matrix

Priority Categorization	Justification
Red = High Priority Finding	This element is categorized as highest priority when: <ul style="list-style-type: none"> <li>• The Data Language(s) provided are incorrect for the dataset.</li> </ul>
Yellow = Medium Priority Finding	Not applicable
Blue = Low Priority Finding	This element is categorized as low priority when: <ul style="list-style-type: none"> <li>• No Data Language is provided when the Data Language is a language other than English (the assumed default language).</li> </ul>
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

Providing a Dataset\_Language is optional. Multiple Dataset\_Languages may be provided, if necessary (Cardinality: 0..\*)

UMM-C Element	Path	Type	Usable Valid Values	Required in DIF 10?	Cardinality	Notes
---------------	------	------	---------------------	---------------------	-------------	-------

DataLanguage	/DIF /Dataset_Language	Enumeration	English	No	0..*	Location of the DatasetLanguageEnum in the DIF 10.3 schema: <a href="https://git.earthdata.nasa.gov/projects/EMFD/repos/dif-schemas/browse/10.x/UmmCommon_1.3.xsd#213">https://git.earthdata.nasa.gov/projects/EMFD/repos/dif-schemas/browse/10.x/UmmCommon_1.3.xsd#213</a>
			Afrikaans			
			Arabic			
			Bosnian			
			Bulgarian			
			Chinese			
			Croatian			
			Czech			
			Danish			
			Dutch			
			Estonian			
			Finnish			
			French			
			German			
			Hebrew			
			Hungarian			
			Indonesian			
			Italian			
			Japanese			
			Korean			
			Latvian			
			Lithuanian			
			Norwegian			
			Polish			
			Portuguese			
			Romanian			
			Russian			
			Slovak			
			Spanish			
			Ukrainian			
			Vietnamese			

### Example Mapping

DIF 10

```
<Dataset_Language>English</Dataset_Language>
```

UMM

```
"DataLanguage": "English"
```

### ECHO 10

Data Language does not map to ECHO 10.

### ISO 19115-2 MENDS

Providing a Data Language is optional. Multiple Data Languages may be provided (Cardinality: 0..\*).

UMM-C Element	Path	Type	Notes
DataLanguage	/gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/ gmd:language/gco:CharacterString  with  /gmi:MI_Metadata/ gmd:identificationInfo/gmd:MD_DataIdentification/ gmd:MD_CharacterSetCode codeList="https://cdn.earthdata.nasa.gov/iso/resources/Codelist/gmxCodelists.xml#MD_CharacterSetCode" codeListValue=UTF8	String  Codelist	Maps to the UMM element Data Language.  Provide the data language (e.g. English) in the Character String (first path listed).  The codelist value of "UTF8" must be provided with the string in order for CMR to properly parse out the data language.

### Example Mapping

#### ISO 19115-2 MENDS

```

<gmi:MI_Metadata>
  ...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      ...
      <gmd:language>
        <gco:CharacterString>English</gco:
CharacterString>
      </gmd:language>
      <gmd:MD_CharacterSetCode codeList="https://cdn.
earthdata.nasa.gov/iso/resources/Codelist
/gmxCodelists.xml#MD_CharacterSetCode"
codeListValue="UTF8">UTF8</gmd:MD_CharacterSetCode>
      ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
  ...
</gmi:MI_Metadata>

```

#### UMM

```
"DataLanguage": "English"
```

### ISO 19115-2 SMAP

Providing a Data Language is optional. Multiple Data Languages may be provided (Cardinality: 0..\*).

UMM-C Element	Path	Type	Notes
DataLanguage	/gmd:DS_Series/gmd:seriesMetadata/gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/ gmd:language/gco:CharacterString  with  /gmd:DS_Series/gmd:seriesMetadata/gmi:MI_Metadata/ gmd:identificationInfo/gmd:MD_DataIdentification/ gmd:MD_CharacterSetCode codeList="https://cdn.earthdata.nasa.gov/iso/resources/Codelist/gmxCodelists.xml#MD_CharacterSetCode" codeListValue=UTF8	String  Codelist	Maps to the UMM element Data Language.  Provide the data language (e.g. English) in the Character String (first path listed).  The codelist value of "UTF8" must be provided with the string in order for CMR to properly parse out the data language.

### Example Mapping

#### ISO 19115-2 SMAP

```

<gmd:DS_Series>
<gmd:seriesMetadata>
<gmi:MI_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      ...
      <gmd:language>
        <gco:CharacterString>English</gco:
CharacterString>
      </gmd:language>
      <gmd:MD_CharacterSetCode codeList="https://cdn.
earthdata.nasa.gov/iso/resources/Codelist
/gmxCodeLists.xml#MD_CharacterSetCode"
codeListValue="UTF8">UTF8</gmd:MD_CharacterSetCode>
    </gmd:MD_DataIdentification>
    ...
  </gmd:identificationInfo>
  ...
</gmi:MI_Metadata>
</gmd:seriesMetadata>
</gmd:DS_Series>

```

UMM

```
"DataLanguage": "English"
```

## UMM Migration

None

## History

### UMM Versioning

Version	Date	What Changed
1.15.5	12/3/2020	No changes were made for Data Language during the transition from version 1.15.4 to 1.15.5
1.15.4	9/18/2020	No changes were made for Data Language during the transition from version 1.15.3 to 1.15.4
1.15.3	7/1/2020	No changes were made for Data Language during the transition from version 1.15.2 to 1.15.3
1.15.2	5/20/2020	No changes were made for Data Language during the transition from version 1.15.1 to 1.15.2
1.15.1	3/25/2020	No changes were made for Data Language during the transition from version 1.15.0 to 1.15.1
1.15.0	2/26/2020	No changes were made for Data Language during the transition from version 1.14.0 to 1.15.0
1.14.0	10/21/2019	No changes were made for Data Language during the transition from version 1.13.0 to 1.14.0
1.13.0	04/11/2019	No changes were made for Data Language during the transition from version 1.12.0 to 1.13.0.
1.12.0	01/22/2019	No changes were made for Data Language during the transition from version 1.11.0 to 1.12.0.
1.11.0	11/28/2018	No changes were made for Data Language during the transition from version 1.10.0 to 1.11.0.
1.10.0	05/02/2018	No changes were made for Data Language during the transition from version 1.9.0 to 1.10.0.

## ARC Documentation

Version	Date	What Changed	Author
1.0	10/12/18	Recommendations/priority matrix transferred from internal ARC documentation to wiki space	<a href="#">Jeanne' le Roux</a>