QA Rule Definitions

From GCMD Wiki

NASA Global Change Master Directory/IDN/CMR

Quality Assessment (QA) Rules and Rule Sets

Document Version: 1.0

(Revision 35672, 2016-2-02)

Content
1. #Introduction
2. #Creating a Rule Set
3. #Rule Type Definitions
4. #Example Rule Set
5. #Using Rule Sets in the QA Tool/QA Viewer

Introduction

This document is designed to guide an interested author in creating a customized Quality Assessment (QA) rule set for use in the QA Tool/QA Viewer applications. These applications are driven by rule sets that assess the quality of a metadata record based on the desired focus and goals of the rule set author. Rule sets are compilations of rules; rules use rule types to test the content provided by a designated xpath within a metadata schema based on a given configuration. Currently, there are 40 rule types of varying complexity:

The simplest rule types require no parameters within their configuration:

RuleType	configuration
FloatPointCheck	

Others require a single parameter, to either (1) configure the rule to check a value or (2) configure the rule to use a list:

	RuleType	configuration
1	MaxFieldLengthCheck	integer is 12000
2	ExactKeywordCheck	check_strings using [CARTESIAN;GEODETIC]

More complicated rule types may combine multiple parameters:

RuleType	configuration
DateOrderCheck	sequence using [Start_Date;Stop_Date] | dateFormat is yyyy-MM-dd

Each specific rule type has its own configuration parameters; to review the rule types and see the list of accompanying parameters, notes on general use, and examples, jump to the #Rule Type Definitions section. The next section - #Creating a Rule Set - provides a general overview of compiling a rule set.

Creating a Rule Set

Rule sets are compilations of rules applied to fields, designated by an xpath, within a metadata schema. Rules can be applied to multiple xpaths, and each xpath can include multiple rules. The rule set author must include at least one rule, but can use as many or as few xpaths and rule types as they feel is necessary to accomplish their quality assessment goals. In addition to setting a rule with an xpath and rule type, the rule set author will also set any applicable configurations, decide on fail severities, and write Fail and/or Pass messages to be displayed in the QA results.

Rule sets are checked into the tool in JSON Format; however, an internal script has been developed to allow rule sets to be written using spreadsheets (i.e, CSV files) The spreadsheets can then be converted to the internal JSON format by GCMD staff and checked into the tool for use.

The following steps are suggested to create a customized rule set using a spreadsheet:

1. Consider the constraints the author would like to impose on their metadata and the desired quality assessment purpose/goals
2. Create a list of the xpaths with their applicable constraints that need to be checked
3. Review the list of Rule Types in the #Rule Type Definitions section and determine which rule types are applicable
4. Assemble the rule set spreadsheet following the formatting and guidance in the #Rule Set Format section below
5. Name the rule set; it is recommended that the name include the Metadata Schema, the Purpose, and the Author; for example, dif_9.9.3-ummc_compliance-gcmd_staff
6. Submit the spreadsheet along with a short description of its purpose to GCMD Staff for review and conversion

Rule set authors interested in writing their rule sets directly in JSON format can contact the GCMD staff for information on formatting requirements.

Rule Set Format

The rule set spreadsheet should be arranged as follows, with each line in the spreadsheet corresponding to a rule:

RuleType	xpath	configuration	Rule Type	severityIfFail	messageIfFail	messageIfPass	Notes/Annotations
Required: predefined	Required: defined by schema	Required: defined by author based on available parameters	Required: predefined	Required: one of ERROR, WARN, or INFO	Optional: author defined	Optional: author defined	Optional: author defined

Where:

RuleType
: The name of the QA Rule Type, which can be found in the #Rule Type Definitions section

xpath
: The xpath within the schema where the author would like to apply the rule

configuration
: The parameters that the rule uses to check the field. See #Rule Type Definitions for information on setting configurations for specific rule types

- General Configuration Conventions:
  - Some rule types do not require a configuration; for these rule types, the configuration is blank
  - For rule types that are setting a value, the configuration uses: <name> is <value>
  - For rule types that are setting a list, the configuration uses: <name> using [a;b;c] where semicolon (';') is used to the separate different strings in the list
  - If a configuration requires multiple parameters, | (Pipe) is used to separate the different parameters
    - Example: <name1> is <value> | <name2> using [a;b;c]

Rule Type
: The name of the Qa Rule Type in Title Case (with a space between each word)

severityIfFail
: The result if the content of the metadata fails the applied rule
  - Possible severity values are:
    - ERROR: The content provided has completely failed the rule type and must be changed to pass quality assurance of the metadata. Authors should set the severity to ERROR if they will not accept values that do not pass the rule.
    - WARN: The content provided did not pass the rule and it is encouraged but not required that the metadata author improve the content of the field. Rule set authors should set rules to WARN if they will accept the given content but would like the metadata author to review and consider changes to improve the quality of the record.
    - INFO: An option to provide additional information about the content of the field.

messageIfFail
: A customizable result message that will be displayed in the QA Tool API or QA Viewer if a field fails a rule

messageIfPass
: A customizable result message that will be displayed in the QA Tool API or QA Viewer if a field passes a rule
  - When writing both Fail Messages and Pass Messages, the following "shortcut" tags can be used:
    - ${xpath} will insert the full xpath (ex: /DIF/Entry_ID) into the message
    - ${name} will insert the Field Name (ex: Entry_ID) into the message
    - ${content} will insert the content used in the field into the message
    - ${size} will insert the total character length in the field into the message
    - ${max} will insert the maximum character length allowed by the configuration parameter for the rule into the message
    - ${min} will insert the minimum character length allowed by the configuration parameter for the rule into the message
    - ${format} will insert the format allowed by the configuration parameter for the rule into the message
    - ${actual} will insert the actual format used in the field into the message

Notes/Annotations
: An optional column for notes about the row; this column will not be included in the conversion

Rule Type Definitions

This section is intended to introduce each specific rule type available for use in rule sets. Each rule type below is accompanied by a short definition, a list of accompanying parameters to use in the "configuration" column of a rule set file, a note on the general use of the rule, and where available, one or more examples using DIF (/DIF/...) or ECHO (/Collection/...) xpaths. Some examples are in use in the current <schema>-ummc_compliance rule sets available in the QA Tool while others are hypothetical.

The names of the Rule Types are intended to indicate their general use or function; however, names may be revised in the future. Rule types may be added (or modified) as user feedback is received. Currently, there are 40 rule types available across six different categories:

1. Character Rule Types, which are applied to the number, type, or pattern of characters allowed within a field
2. Controlled Vocabulary Rule Types, which check that the content of the field matches a valid keyword, either by comparing it to an author-provided list or to an external source, such as KMS
3. Date Rule Types, which are applied to date fields
4. Link Rule Types, which are applied to fields that are or may contain links to an external source in the metadata
5. Miscellaneous Rule Types, which contain various "other" rule types
6. Numeric Rule Types, which are applied to fields where the content should be or should include a numeric value

In addition, the #Optional Parameters section describes additional parameters that can be added to a configuration but are independent of any specific rule type

To review the rule types in detail, select a Rule Category or individual name from the table below, or scroll through the full list (alphabetical by category)

Rule Name	Rule Category
#AllURLsExistCheck	#Link Rule Types
#URLFormatCheck	Link Rule Types
#DoiExistsCheck	Link Rule Types
#DoiFormatCheck	Link Rule Types
#SingleURLExistsCheck	Link Rule Types
#ARKFormatCheck	Link Rule Types
#ARKExistsCheck	Link Rule Types
#CharacterPatternAllowed	#Character Rule Types
#CharacterPatternDenied	Character Rule Types
#MaxFieldLengthCheck	Character Rule Types
#MinFieldLengthCheck	Character Rule Types
#RequiredCharacterCheck	Character Rule Types
#RestrictedCharacterCheck	Character Rule Types
#DateFormatCheck	#Date Rule Types
#DateOrderCheck	Date Rule Types
#MultiDateFormatsCheck	Date Rule Types
#FloatPointCheck	#Numeric Rule Types
#FloatRangeCheck	Numeric Rule Types
#IntegerPointCheck	Numeric Rule Types
#FloatWithUnitInsideFieldCheck	Numeric Rule Types
#FloatWithUnitAtFieldEndCheck	Numeric Rule Types
#FloatWithUnitAtFieldStartCheck	Numeric Rule Types
#IntegerWithUnitInsideField	Numeric Rule Types
#IntegerWithUnitAtFieldEndCheck	Numeric Rule Types
#IntegerWithUnitAtFieldStartCheck	Numeric Rule Types
#ControlledKeywordCheck	#Controlled Vocabulary Rule Types
#DuplicatedKeywordCheck	Controlled Vocabulary Rule Types
ISOControlledKeywordCheck	Controlled Vocabulary Rule Types
#ExactKeywordCheck	Controlled Vocabulary Rule Types
#ContainsCheck	#Miscellaneous Rule Types
#EntryIDExistsCheck	Miscellaneous Rule Types
#FieldExistsCheck	Miscellaneous Rule Types
#IdenticalFieldCheck	Miscellaneous Rule Types
#FieldShouldNotExistCheck	Miscellaneous Rule Types
#GranulesSpatialCoordinatesCheck	Miscellaneous Rule Types
#NonidenticalFieldCheck	Miscellaneous Rule Types
#PrefixCheck	Miscellaneous Rule Types
#RequiredFieldCheck	Miscellaneous Rule Types
#RequiredChoiceFieldCheck	Miscellaneous Rule Types
#SpatialCoordinatesCheck	Miscellaneous Rule Types
#SuffixCheck	Miscellaneous Rule Types
#UniqueEntryIDCheck	Miscellaneous Rule Types
#WhiteSpaceCheck	Miscellaneous Rule Types
#Conditional Check	#Optional Parameters
#Case-Sensitivity	Optional Parameters

Character Rule Types

Character rule types are applied to the number, type, or pattern of characters that are allowed within a field.

CharacterPatternAllowed

- Definition: this rule type checks that a complex pattern of characters defined by a Regular expression is used in the field.
- Use with: fields where the author would like to control the pattern of characters in the content but cannot do so using other existing rule types in this document.
- Parameter: <regular-expression is > a regular expression "[a-z1-9]"

Example:

1. Check that Processing/Product Level follows the correct pattern of Numbers 1-4 and Letters a or b.

	RuleType	xpath	Configuration
1	CharacterPatternAllowed	/DIF/ProductLevelID	regular-expression is [0-4]{1}[a-b]?
2	CharacterPatternAllowed	/Collection/ProcessingLevelId	regular-expression is [0-4]{1}[a-b]?

These examples are hypothetical uses of this rule type

CharacterPatternDenied

- Definition: this rule type checks that a complex pattern of characters defined by a Regular expression is not used in the field.
- Use with: fields where the author would like to control the pattern of characters in the content but cannot do so using other existing rule types in this document.
- Parameter: <regular-expression is> a regular expression "[a-z1-9]"

MaxFieldLengthCheck

- Definition: this rule type checks that the length of the field is within the maximum length allowed for that field.
- Parameter: <integer is> the maximum length of the field

Example:

1. The DIF Abstract should be no longer than 12,000 characters to be UMMC-compliant
2. The ECHO Description should be no longer than 12,000 characters to be UMMC-compliant

	RuleType	xpath	Configuration
1	MaxFieldLengthCheck	/DIF/Summary/Abstract	integer is 12000
2	MaxFieldLengthCheck	/Collection/Description	integer is 12000

MinFieldLengthCheck

- Definition: this rule type checks that the length of the field is greater than a designated minimum length.
- Use with: any field that should have a minimum length.
- Parameter: <integer is> the minimum length of the field

Example:

1. Set the minimum length of the DIF Entry ID to 5 characters
2. Set the minimum length of the ECHO Short Name to 5 characters

	RuleType	xpath	Configuration
1	MinFieldLengthCheck	/DIF/Entry_ID	integer is 5
2	MinFieldLengthCheck	/Collection/ShortName	integer is 5

These examples are hypothetical uses of this rule type

RequiredCharacterCheck

- Definition: this rule type checks that a required character is present in the field.
- Use with: any field where the author would like to require the use of a specific character.
- Parameter: <characters is> any character

Example:

1. Check the author includes the "@" character in an email address field.

	RuleType	xpath	Configuration
1	RequiredCharacterCheck	/DIF/Personnel/Email	characters is @
2	RequiredCharacterCheck	/Collection/Contacts/Contact/OrganizationEmails/Email	characters is @

These examples are hypothetical uses of this rule type

RestrictedCharacterCheck

- Definition: this rule type checks that restricted characters are not present in the field.
- Use with: any field where the author would like to restrict the use of a specific character.
- Parameter: <characters is> any character

Example:

1. Do not allow Entry IDs to include \, /, or : characters.
2. Do not allow Short Names to include \, /, or : characters.

	RuleType	xpath	Configuration
1	RestrictedCharacterCheck	/DIF/Entry_ID	characters is \\/:
2	RestrictedCharacterCheck	/Collection/ShortName	characters is \\/:

These examples are hypothetical uses of this rule type

back to table of contents

Controlled Vocabulary Rule Types

These rule types will check that the content of the field matches a valid keyword, either by comparing to an author-provided list or to an external source, such as KMS.

ControlledKeywordCheck

- Definition: this rule type checks that the keyword provided comes from a predefined, controlled list, such as KMS.
- Use with: any field that should be populated with a controlled keyword that is defined by KMS.
  - If the keyword is comprised of multiple parts (such as a Short Name & Long Name), the rule type should be applied to the parent xpath.
- Parameter: none (if the rule type should apply to a single field or, if subfields are present, all subfields are to checked against KMS)
- Parameter (optional): <ignore using> [field name(s) to ignore]
- Parameter (optional): <include using> [field name(s) to include]

Example:

1. Check that the metadata author provided a valid Data Center. All subfields should be checked.
2. Check that the metadata author provided a valid Science Keyword. The Detailed Variable subfield is free-text and should be ignored.
3. Check that the metadata author provided a valid Platform. The check should only look at the Short_Name & Long_Name subfields.

	RuleType	xpath	configuration
1	ControlledKeywordCheck	/DIF/Data_Center/Data_Center_Name	
2	ControlledKeywordCheck	/DIF/Parameters	ignore using [Detailed_Variable]
3	ControlledKeywordCheck	/DIF/Platform	include using [Short_Name;Long_Name]

DuplicatedKeywordCheck

- Definition: this rule type checks if there are duplicate keywords for a given keyword category
- Use with: any keyword field
- Parameter: none (if the rule type should apply to a single field or, if subfields are present, all subfields[
- Parameter (optional): <ignore using> [field name(s) to ignore]
- Parameter (optional): <include using> [field name(s) to include]

ExactKeywordCheck

- Definition: this rule type checks that the text entered into the field exactly matches a value defined by the rule, such as an enumerations list.
- Use with: any field that should be populated with a controlled keyword that is defined by the rule author.
- Parameter: <check_strings using> [a list of acceptable strings]

Example:

1. Check that an accepted Coordinate System is provided. The metadata author should only use CARTESIAN or GEODETIC.
2. Check that an accepted Contact Role is provided. The metadata author should only use INVESTIGATOR, TECHNICAL CONTACT, or METADATA AUTHOR.

	RuleType	xpath	configuration
1	ExactKeywordCheck	/Collection /Spatial /HorizontalSpatialDomain /Geometry /CoordinateSystem	check_strings using [CARTESIAN;GEODETIC]
2	ExactKeywordCheck	/DIF/Personnel/Role	check_strings using [INVESTIGATOR;TECHNICAL CONTACT;METADATA AUTHOR]

ISOControlledKeywordCheck

- Definition: this rule type checks that the keyword in an ISO metadata record provided comes from a predefined, controlled list, such as KMS.
- Use with: any field that should be populated with a controlled keyword that is defined by KMS.
- Parameter: <keyword_scheme> is one of: science_keyword, project, platform, instrument, sensor, or location
- NOTE: The check currently treats all keywords as a block; if one keyword fails among many, the check will fail but does not indicate which keyword should be corrected

back to table of contents

Date Rule Types

Date rule types apply to Date Fields.

- Note: there is java formatting required for more complicated dateTime formats. See Class SimpleDateFormat (http://www.example.com) for examples

DateFormatCheck

- Definition: this rule type checks that the date provided follows a given date format
- Use with: any date field where only one date format is allowed
  - To allow multiple date formats and/or enumerations, see #MultiDateFormatsCheck
- Parameter: <Date is> a date format

Example:

1. In DIF 9, the Creation Date should follow the "yyy-MM-dd" format.
2. IN ECHO, the Single DateTime should follow the "yyyy-MM-ddTHH:mm:ssZ" format. The java required syntax that should be used in the configuration is yyyy-MM-dd'T'HH:mm:ssX

	RuleType	xpath	Configuration
1	DateFormatCheck	/DIF/DIF_Creation_Date	Date is yyyy-MM-dd
2	DateFormatCheck	/Collection/Temporal/SingleDateTime	Date is yyyy-MM-dd'T'HH:mm:ssX

DateOrderCheck

- Definition: this rule type checks that dates are in chronological order, with start dates occurring before stop dates.
- Use with: paired date fields that should be in a chronological order.
- Parameter: <sequence using> [the correct order of the date fields]
- Parameter: <dateFormat is> a date format

Example:

1. In DIF, check that the Temporal Coverage Start date occurs before the stop date, using the "yyyy-MM-dd" format.
2. IN ECHO, check that the Temporal RangeDateTimes occur in chronological order, using the "yyyy-mm-ddTHH:MM:SSZ" format. The required java syntax is yyyy-MM-dd'T'HH:mm:ssX

	RuleType	xpath	Configuration
1	DateOrderCheck	/DIF/Temporal_Coverage	sequence using [Start_Date;Stop_Date] | dateFormat is yyyy-MM-dd
2	DateOrderCheck	/Collection/Temporal /RangeDateTime/	sequence using [BeginningDateTime;EndingDateTime] | dateFormat is yyyy-MM-dd'T'HH:mm:ssX

MultiDateFormatsCheck

- Definition: this rule type checks that an entry in a date field follows one of multiple allowed formats or uses a specific keyword
- Use with: date fields where multiple date formats or enumerations/keywords are allowed
- Parameter: <date_formats using> [dateFormat1;dateFormat2]
- Parameter: <check_strings using> [Keyword1;Keyword2]

Example:

1. In DIF 10, Metadata Creation should be a date, dateTime, or one of: Not provided, unknown, present, or unbounded
2. In ECHO 10, check that the dateTime used in the RangeDateTime field is either yyyy-MM-ddTHH:MM:SSZ or yyyy-MM-ddTHH:MM:SS.sssZ

	RuleType	xpath	Configuration
1	MultiDateFormatsCheck	/DIF/Metadata_Dates /Metadata_Creation	date_formats using [yyyy-MM-dd;yyyy-MM-dd'T'HH:mm:ssX] | check_strings using [Not provided;unknown;present;unbounded]
2	MultiDateFormatsCheck	/Collection/Temporal/RangeDateTime /BeginningDateTime	date_formats using [yyyy-MM-dd'T'HH:mm:ss.SSSX;yyyy-MM-dd'T'HH:mm:ssX]

back to table of contents

Link Rule Types

These rule types are applied to fields that are or may contain links to an external source in the metadata.

AllURLsExistCheck

- Definition: this rule type checks that all URLs within a text field resolve to an existing URL if the URL starts with http://, https:// or ftp://
  - There may be false-positives associated with this rule type if the URL takes too long to resolve
- Use with fields that may contain URLs, such as Abstracts, Purposes, Use Constraints, etc.
- Parameter: none

Example:

1. Check that any URLs within a DIF Abstract are valid. There may be false positives, so the rule severity is set to WARN.
2. Check that any URLs within an ECHO Description are valid. There may be false positives, so the rule severity is set to WARN.

	RuleType	xpath	Configuration
1	AllURLsExistCheck	/DIF/Summary/Abstract	
2	AllURLsExistCheck	/Collection/Description	

ARKExistsCheck

- Definition: this rule type checks than an ARK identifier used in a field exists
- Use with: any field that should be an ARK identifier
- Parameter: none

Example:

1. Verify that an ARK used in the DIF 10 Dataset Citation exists
2. Verify that an ARK used in the DIF 10 Reference exists

	RuleType	xpath	configuration
1.	ARKExistsCheck	/DIF/Dataset_Citation/Persistent_Identifier/Identifier	
2.	ARKExistsCheck	/DIF/Reference/Persistent_Identifier/Identifier	

ARKFormatCheck

- Definition: this rule type checks that an ARK is properly formatted.
- Use with: any field that should be an ARK identifier
- Parameter: none

Example:

1. Verify that an ARK identifier used in the DIF 10 Dataset Citation is properly formatted.
2. Verify that an ARK identifier used in the DIF 10 Reference is properly formatted.

	RuleType	xpath	configuration
1.	ARKFormatCheck	/DIF/Dataset_Citation/Persistent_Identifier/Identifier	
2.	ARKFormatCheck	/DIF/Reference/Persistent_Identifier/Identifier	

DoiExistsCheck

- Definition: this rule type checks that a DOI used in a field exists
- Use with: any field that should be a DOI.
- Parameter: none

Example:

1. Verify that a DOI used in the DIF Data Set Citation resolves to an existing URL.
2. Verify that a DOI used in a DIF Reference resolves to an existing URL.

	RuleType	xpath	Configuration
1	DoiExistsCheck	/DIF/Data_Set_Citation/Dataset_DOI	
2	DoiExistsCheck	/DIF/Reference/DOI	

DoiFormatCheck

- Definition: this rule type checks that a DOI is properly formatted.
- Use with: any field that should be a DOI.
- Parameter: none

Example:

1. Verify that a DOI used in the DIF Data Set Citation is properly formatted.
2. Verify that a DOI used in a DIF Reference is properly formatted.

	RuleType	xpath	Configuration
1	DoiFormatCheck	/DIF/Data_Set_Citation/Dataset_DOI	
2	DoiFormatCheck	/DIF/Reference/DOI	

SingleURLExistsCheck

- Definition: this rule type checks that the URL in a URL-only field resolves to an existing URL
  - There may be false-positives associated with this rule if the URL takes too long to resolve
- Use with: any field that is a URL.
- Parameter: none

Example:

1. Check that a DIF Related URL resolves to an existing URL. There may be false positives, so the rule severity is set to WARN.
2. Check that a ECHO Online Resource URL resolves to an existing URL. There may be false positives, so the rule severity is set to WARN.

	RuleType	xpath	Configuration
1	SingleURLExistsCheck	/DIF/Related_URL/URL	
2	URLSingleURLExistsCheck	/Collection/OnlineResources/OnlineResource/	

URLFormatCheck

- Definition: this rule type checks that a URL is properly formatted.
- Use with: any URL field.
- Parameter: none

Example:

1. Check for valid formatting in the DIF Related URL.
2. Check for valid formatting in the ECHO Online Resource.

	RuleType	xpath	Configuration
1	URLFormatCheck	/DIF/Related_URL/URL	
2	URLFormatCheck	/Collection/OnlineResources/OnlineResource/URL	

back to table of contents

Miscellaneous Rule Types

ContainsCheck

- Definition: this rule type checks that a specific set variable is present somewhere in the field.
- Use with: any field where the author wants to check that certain words or other values occur somewhere within the field
- Parameter: <check_strings using> [one or more strings of text]
- Parameter: <check_type is> one of "require-one", "require-all", "require-some" or "not-allowed"

Example:

1. A DAAC may want the name of the Platform and Instrument used to be included in the Entry Title.
2. A DAAC may want their DAAC acronym to be included in the Entry ID.

	RuleType	xpath	Configuration
1	ContainsCheck	/DIF/Entry_Title	check_strings using [MODIS_AQUA; MODIS_TERRA] | check_type is require-all
2	ContainsCheck	/DIF/Entry_ID	check_strings using [GES_DISC] | check_type is require-all

These examples are hypothetical uses of this rule type

EntryIDExistsCheck

- Definition: this rule type checks that an EntryID used currently exists. This is distinct from the #UniqueEntryIDCheck.
- Use with: any field that references a DIF Entry ID.
- Parameter: none

Example:

1. Check that the Parent DIF supplied in a DIF record exists.
2. Check that the Associated DIF supplied in an ECHO record exists.

	RuleType	xpath	Configuration
1	EntryIDExistsCheck	/DIF/Parent_DIF	
2	EntryIDExistsCheck	/Collection/AssociatedDIFs/DIF/EntryId	

FieldExistsCheck

- Definition: this rule type checks that content is provided in the field.
- Use with: any field where the author would like to check for content if the xpath to the field is present in the metadata record being checked.
- Parameter: none

Example:

1. Check that Ancillary Keywords are populated in the DIF
2. Check that TwoDCoordinateSystems Name is populated in ECHO.

	RuleType	xpath	Configuration
1	FieldExistsCheck	/DIF/Keyword	
2	FieldExistsCheck	/Collection/TwoDCoordinateSystems/TwoDCoordinateSystem/TwoDCoordinateSystemName	

FieldShouldNotExistCheck

- Definition: This rule will check for content in a field that should not exist.
- Use with: Any field that the author would like not to be populated.
- Parameter: None

Example:

1. Discourage the use of Discipline Name, a deprecated field in the DIF

	RuleType	xpath	configuration
1.	FieldShouldNotExistCheck	/DIF/Discipline/Discipline_Name	

GranulesSpatialCoordinatesCheck

- Definition: this rule will check that the spatial coordinates of any granules associated with a DIF or ECHO record are within the spatial boundary of that DIF or ECHO record
- Use with: any DIF or ECHO record
- Parameter: <west is>the field name that corresponds to the most western coordinate
- Parameter: <east is> the field name that corresponds to the most eastern coordinate
- Parameter: <south is> the field name that corresponds to the most southern coordinate
- Parameter: <north is> the field name that corresponds to the most northern coordinate

- Parameter: <record_id is> the Collection ID
  - The following query can be used to locate the Collection ID, if not known: https://cmr.earthdata.nasa.gov/search/collections?dif_entry_id=[DIF Entry_ID]

Example:

1. Check that any existing granules are within the spatial bounds of the DIF record "CD15_PRODUCTIVITY"
2. Check that any existing granules are within the spatial bounds of the ECHO record "MYD17A2, Version 5"

	RuleType	xpath	configuration
1.	GranulesSpatialCoordinatesCheck	/DIF/Spatial_Coverage	east is Easternmost_Longitude | north is Northernmost_Latitude | south is Southernmost_Latitude | record_id is C179124905-ORNL_DAAC
2.	GranulesSpatialCoordinatesCheck	/Collection/Spatial /HorizontalSpatialDomain/Geometry /BoundingRectangle	west is WestBoundingCoordinate | east is EastBoundingCoordinate | north is NorthBoundingCoordinate | south is SouthBoundingCoordinate | record_id is C117500875-LPDAAC_ECS

These examples are hypothetical uses of this rule type

IdenticalFieldCheck

- Definition: this rule type checks if the content of two distinct fields contains the same values
- Use with: any two fields that should have identical content
- Parameter: <other is > the name of the field to be compared

Example:

1. The author may want the Entry Title to match the Dataset Title in the DIF Citation field.
2. The author may want the Collection Short Name to match the listed Associated DIF Entry ID

	RuleType	xpath	Configuration
1	IdenticalFieldCheck	/DIF/Data_Set_Citation/Dataset_Title	other is Entry_Title
2	IdenticalFieldCheck	/Collection/ShortName	other is EntryID

These examples are hypothetical uses of this rule type

NonidenticalFieldCheck

- Definition: this rule type checks if the content of two distinct fields contains different values
- Use with: any two fields where the content should not match
- Parameter: <other is> the name of the field to be compared

Example:

1. Check that the Entry ID is different from the Entry Title, to encourage more descriptive titles.
2. Check that the Collection Description is different from the DataSetID, to encourage a more descriptive description

	RuleType	xpath	Configuration
1	NonidenticalFieldCheck	/DIF/Entry_Title	other is Entry_ID
2	NonidenticalFieldCheck	/Collection/Description	other is DataSetID

These examples are hypothetical uses of this rule type

PrefixCheck

- Definition: this rule type checks that the prefix used in the field matches values allowed by the rule
- Use with: fields that should start with a set prefix
- Parameter: <prefixes using> [a list of prefixes]

Example:

1. If the Entry ID or Short Name for a metadata record should be a doi, use the Prefix check to require that the ID starts with "doi:"

	RuleType	xpath	Configuration
1	PrefixCheck	/DIF/Entry_ID	prefixes using [doi:]
2	PrefixCheck	/Collection/ShortName	prefixes using [doi:]

These examples are hypothetical uses of this rule type

RequiredFieldCheck

- Definition: this rule type checks for the presence of a field in the metadata and can be used to check for fields above and beyond those required by the metadata schema.
- Use with: any field that the rule author would like to require.
- Parameter: <field is> the xpath of the field

Example:

1. The UMM-C requires several DIF fields that are not required by the DIF 9.9.3 schema, such as Platform. Check that the metadata author provided a Platform in their DIF.

2. Granule Spatial Representation is required by the UMM-C. Check that the metadata author provided a Granule Spatial Representation in their ECHO record.

	RuleType	xpath	configuration
1	RequiredFieldCheck	/DIF/Source	field is /DIF/Source
2	RequiredFieldCheck	/Collection/Spatial/GranuleSpatialRepresentation	field is /Collection/Spatial/GranuleSpatialRepresentation

RequiredChoiceFieldCheck

- Definition: this rule type checks that one of several required options is used in the metadata record
- Use with: any field that has subfields where one of several is required, such as Temporal_Coverage or Spatial_Coverage
- Parameter: <choice_field> is the xpath of the parent field
- Parameter: <field_list> using [the name of the subfields, of which one is required to be used]

Example:

1. The DIF10 schema requires one of the following to be filled in under Temporal Coverage: Range_DateTime; Single_DateTime; Periodic_DateTime; or Paleo_DateTime for compliance with the UMM-C
2. The DIF10 schema allows a choice of Contact_Person or Contact_Group under Organization/Personnel, one of which is required.

	RuleType	xpath	configuration
1	RequiredChoiceFieldCheck	/DIF/Temporal_Coverage	field_list using [Range_DateTime;Single_DateTime;Periodic_DateTime;Paleo_DateTime]
2	RequiredChoiceFieldCheck	/DIF/Organization /Personnel	field_list using [Contact_Person;Contact_Group]

SpatialCoordinatesCheck

- Definition: this rule type checks that spatial coordinates are valid.
- Use with: spatial bounding box fields.
- Parameter: <west is>the field name that corresponds to the most western coordinate
- Parameter: <east is> the field name that corresponds to the most eastern coordinate
- Parameter: <north is> the field name that corresponds to the most northern coordinate
- Parameter: <south is> the field name that corresponds to the most southern coordinate

Example:

1. Check that all values are valid in the DIF Spatial Coverage fields.
2. Check that all values are valid in the ECHO Bounding Rectangle fields.

	RuleType	xpath	Configuration
1	SpatialCoordinatesCheck	/DIF/Spatial_Coverage	west is Westernmost_Longitude | east is Easternmost_Longitude | north is Northernmost_Latitude | south is Southernmost_Latitude
2	SpatialCoordinatesCheck	/Collection/Spatial/HorizontalSpatialDomain /Geometry/BoundingRectangle	west is WestBoundingCoordinate | east is EastBoundingCoordinate | north is NorthBoundingCoordinate | south is SouthBoundingCoordinate

SuffixCheck

- Definition: this rule type checks that the suffix used in the field matches values allowed by the rule set.
- Use with: fields that should end with a set suffix.
- Parameter: <postfixes using> [a list of suffixes]

Example:

1. Check that the Maximum Depth value in DIF Spatial Coverage is entered as meters, kilometers, feet, surface, hecto Pascals, or millibars
2. Check that the Swath Width of ECHO Spatial Orbit Parameters is entered as meters or kilometers.

	RuleType	xpath	Configuration
1	SuffixCheck	/DIF/Spatial_Coverage/Maximum_Depth	postfixes using [m;meters;km;kilometers;ft;feet;surface;hPA;hecto Pascals;mb;millibars]
2	SuffixCheck	/Collection/Spatial/OrbitParameters /SwathWidth	postfixes using [m;km]

UniqueEntryIDCheck

- Definition: this rule type checks that the EntryID used is a unique value. This is distinct from #EntryIDExistsCheck.
- Use with: the EntryID field to check if the ID entered already exists in the GCMD.
- Parameter: none

Example:

1. Check if the EntryID of a record is already in use. EntryIDs for updated records will already exist, so the rule severity should only be set to WARN, to bring the EntryID to the attention of the user.

	RuleType	xpath	Configuration
1	UniqueEntryIDCheck	/DIF/Entry_ID	

WhiteSpaceCheck

- Definition: this rule type checks for the presence of excess white space within a text field
- Use with: any text field, such as Abstract or Quality, where there may be issues with excess white space. Depending on the parameter used below, the rule will fail if excess white space is found in the field
- Parameter: <ActOnTrailingSpaces is> threshold number of trailing spaces that should fail the rule
  - Default: will fail on any trailing spaces
- Parameter: <ActOnLeadingSpaces is> threshold number of leading spaces that should fail the rule
  - Default: will not fail on any leading spaces
- Parameter: <ActOnConsecutiveNonNewLineWhiteSpaces is> threshold number of consecutive white spaces that should fail the rule
  - Default: will fail on more than 2 consecutive spaces in content, unless after a period
- Parameter: <ActOnConsecutiveNewLines is> threshold number of consecutive new lines that should fail the rule
  - Default: will fail on more than 2 lines in a row
- Parameter: <ActOnSpaceAfterPeriod is> threshold number of white spaces after a period that should fail the rule

Example:

1. Check for excess white space in the DIF Abstract using the default settings
2. Allow up to five leading spaces and three consecutive new lines in the ECHO Description

	RuleType	xpath	configuration
1.	WhiteSpaceCheck	/DIF/Abstract	
2.	WhiteSpaceCheck	/Collection/Description	ActOnLeadingSpaces is 5 | ActOnConsecutiveNewLines is 4

These examples are hypothetical uses of this rule type

back to table of contents

Numeric Rule Types

These rule types are applied to fields where the content should be or include a numeric value.

FloatPointCheck

- Definition: this rule type checks that values provided in a given field are decimal values.
- Use with: any number field that should contain a decimal value.
- Parameter: none

Example:

1. Check that the Latitude Resolution within DIF Data Resolution is a decimal value.
2. Check that a Point Longitude in ECHO is a decimal value.

	RuleType	xpath	Configuration
1	FloatPointCheck	/DIF/Data_Resolution/Latitude_Resolution	
2	FloatPointCheck	/Collection/Spatial/HorizontalSpatialDomain/Geometry/Point/PointLongitude	

FloatRangeCheck

- Definition: This rule type checks that a number in decimal format is within a designated range.
- Use with: any number fields that should contain decimal values within a designated range.
- Parameter: <min is> the minimum value
- Parameter: <max is> the maximum value

Example:

1. Check that a DIF Latitude value is between -90 and 90 degrees.
2. Check that a Point Longitude in ECHO is with -180 and 180 degrees.

	RuleType	xpath	Configuration
1	FloatRangeCheck	/DIF/Spatial_Coverage/Northernmost_Latitude	min is -90 | max is 90
2	FloatRangeCheck	/Collection/Spatial/HorizontalSpatialDomain/Geometry/Point/PointLongitude	min is -180 | max is 180

FloatWithUnitAtFieldEndCheck

- Definition: this rule type checks that a field begins with a decimal value followed by a unit
- Use with: any text field
- Parameter: <check_strings using> [unit1;unit2]

FloatWithUnitAtFieldStartCheck

- Definition: this rule type checks that a field ends with a decimal value followed by a unit
- Use with: any text field
- Parameter: <check_strings using> [unit1;unit2]

FloatWithUnitInsideFieldCheck

- Definition: this rule type checks that a decimal value followed by a unit occurs somewhere within a field

- Use with: any text field
- Parameter: <check_strings using> [unit1;unit2]

Example:

1. Check that the Maximum Depth value in DIF Spatial Coverage is a decimal value followed by: meters, kilometers, feet, surface, hecto Pascals, or millibars
2. Check that the Swath Width of ECHO Spatial Orbit Parameters is a decimal value followed by "m" or "km"

	RuleType	xpath	Configuration
1	FloatWithUnitInsideFieldCheck	/DIF/Spatial_Coverage /Maximum_Depth	check_strings using [m;meters;km;kilometers;ft;feet;surface;hPA;hecto Pascals;mb;millibars]
2	FloatWithUnitInsideFieldCheck	/Collection/Spatial/OrbitParameters /SwathWidth	check_strings using [m;km]

IntegerPointCheck

- Definition: this rule type checks that a value provided in a field is an integer
- Use with: any number field that should be an integer.
- Parameter: none

Example:

1. For Number of Sensors of an Instrument, check that the number provided is an integer.
2. For ECHO Number of Orbits, check that the number provided is an integer.

	RuleType	xpath	Configuration
1	IntegerPointCheck	/DIF/Platform/Instrument/NumberOfSensors	
2	IntegerCheck	/Collection/Spatial/OrbitParameters/NumberofOrbits	

These examples are hypothetical uses of this rule type

IntegerWithUnitAtFieldEndCheck

- Definition: this rule type checks that a field ends with an integer value followed by a unit
- Use with: any text field
- Parameter: <check_strings using> [unit1;unit2]

IntegerWithUnitAtFieldStartCheck

- Definition: this rule type checks that a field begins with an integer value followed by a unit
- Use with: any text field
- Parameter: <check_strings using> [unit1;unit2]

IntegerWithUnitInsideFieldCheck

- Definition: this rule type checks that an integer value followed by a unit occurs somewhere within a field
- Use with: any text field
- Parameter: <check_strings using> [unit1;unit2]

Example:

1. Check that the Citation field in the DIF 10 Reference includes an integer value followed by "pp."

	RuleType	xpath	configuration
1.	IntegerWithUnitInsideFieldCheck	/DIF/Reference/Citation	check_strings using [pp.]

back to table of contents

Optional Parameters

There are optional parameters that can be added to the configuration of one or more of the rule types above. Following the parameters required for the particular rule type, insert a | and add the optional parameter that should be used

Conditional Check

- Definition: this optional parameter will allow the rule set author to designate certain rule types to run only if a given condition also exists in the metadata record
- Use with: any rule type above
- Parameter: <conditional_field is> the xpath to the field that should exist to initiate the rule
- Parameter: <conditional_field_content is> the particular content that should exist in the conditional field to initiate the rule

Example:

1. Set the maximum length of an Ancillary Keyword to 160 characters when the Originating Metadata Node of the DIF is GCMD
2. Use the DoiFormatCheck when the Persistent Identifier Type in the DIF 10 is "DOI"

	RuleType	xpath	configuration
1.	MaxFieldLengthCheck	/DIF/Keyword	integer is 160 | condition_field is /DIF/Originating_Metadata_Node | condition_field_content is GCMD
2.	DoiFormatCheck	/DIF/Persistent_Identifier /Identifier	condition_field is /DIF/Persistent_Identifier/Type | condition_field_content is DOI

These examples are hypothetical uses of this rule type

Case-Sensitivity

- Definition: this optional parameter will tell the rule to ignore case when used with rule types that reference a specific value or list of values. By default, all content checks are case-sensitive
- Use with: any rule that is checking a specific value or list of values, such #ExactKeywordCheck or #SuffixCheck
- Parameter: case is insignificant

Example:

1. Use the ExactKeywordCheck to set an approved list of values for Distribution Media but ignore case
2. Use the SuffixCheck to set an approved list of units for Distribution Size but ignore case

	RuleType	xpath	configuration
1.	ExactKeywordCheck	/DIF/Distribution/Distribution_Media	check_strings using [HTTP;FTP] | case is insignificant
2.	SuffixCheck	/DIF/Distribution/Distribution_Size	postfixes using [MB, KB] | case is insignificant

These examples are hypothetical uses of this rule type

back to table of contents

Example Rule Set

The table below is a hypothetical demonstration of a rule set as it should be completed in a spreadsheet.

RuleType	xpath	configuration	Rule Type	severityIfFail	messageIfFail	messageIfPass
AllURLsExistCheck	/DIF/Summary/Abstract		ERROR	All URLs Exist Check	Potential broken link in ${name}	All links in ${name} are valid
ContainsCheck	/DIF/Entry_Title	check_type is require-one	Contains Check	ERROR	Your ${name} does not include one of the required strings. Please include either "MODIS_AQUA" or "MODIS_TERRA" in your title	
ControlledKeywordCheck	/DIF/Parameters	ignore using [Detailed_Variable]	Controlled Keyword Check	ERROR	${content} is not a valid keyword. Please use a valid GCMD Science keyword.	${content} is a valid GCMD Science Keyword
DateFormatCheck	/DIF/DIF_Creation_Date	Date is yyyy-MM-dd	Date Format Check	ERROR	${name} should be in ${format} date format only. Your ${name} is formatted ${actual}	${name} is formatted properly
DateOrderCheck	/DIF/Temporal_Coverage	dateFormat is yyyy-MM-dd	Date Order Check	ERROR	The ${name} Start Date must occur before the Stop Date	${name} is in correct chronological order
DoiExistsCheck	/DIF/Data_Set_Citation /Dataset_DOI		Doi Exists Check	ERROR	Potential broken DOI found in ${name}	${name} is a valid DOI
DoiFormatCheck	/DIF/Data_Set_Citation /Dataset_DOI		Doi Format Check	ERROR	DOI is not formatted properly in ${name}	DOI is formatted properly in ${name}
EntryIDExistsCheck	/DIF/Parent_Dif		EntryID Exists Check	ERROR	${content} is not an existing Entry ID. Review the Entry ID provided	${content} is an existing Entry ID
ExactKeywordCheck	/DIF/Personnel/Role	check_strings using [INVESTIGATOR;TECHNICAL CONTACT;DIF AUTHOR]	Exact Keyword Check	ERROR	${content} is not a valid ${name}. Use one of the following: INVESTIGATOR, TECHNICAL CONTACT, DIF AUTHOR	${content} is a valid ${name}
FieldExistsCheck	/DIF/Keyword		Field Exists Check	ERROR	${name} should be included	${name} is included
FloatPointCheck	/DIF/Data_Set_Citation /Version		Float Point Check	ERROR	Use a decimal value for ${xpath}	
FloatRangeCheck	/DIF/Spatial_Coverage /Northernmost_Latitude	max is 90.0	Float Range Check	ERROR	${name} should be a decimal value between ${min} and ${max}	
IdenticalFieldCheck	/DIF/Entry_Title	other is Data_Set_Title	Identical Field Check	ERROR	The Entry ID and Dataset Title should match. Review these fields	
IntegerPointCheck	/DIF/Data_Set_Citation /Issue_Identification		Integer Point Check	ERROR	Use an integer value for ${xpath}	
MaxFieldLengthCheck	/DIF/Summary/Purpose	integer is 100	Max Field Length Check	ERROR	${name} length should be greater than ${min} characters. Your ${name} is only ${size}.	${name} is ${size} characters
MinFieldLengthCheck	/DIF/Entry_ID	integer is 5	Min Field Length Check	ERROR	${name} length should be greater than ${min} characters. Your ${name} is only	${name} is ${size} characters

					${size}.	
NonidenticalFieldCheck	/DIF/Entry_Title	other is Entry_ID	Nonidentical Field Check	ERROR	Entry Title matches Entry ID. Consider a more descriptive title	
PrefixCheck	/DIF/Data_Set_Citation /Dataset_Release_Place	prefixes using [NASA]	Prefix Check	ERROR	${name} must start with "NASA"	
RequiredCharacterCheck	/DIF/Personnel/Email	character is @	Required Character Check	ERROR	${email} should include the @ character	
RequiredFieldCheck	/DIF/Source	field is /DIF/Source	Required Field Check	ERROR	${name} is required. Please add platform information	
RestrictedCharacterCheck	/DIF/Entry_ID	character is :	Restricted Character Check	ERROR	${name} should not contain ":"	
SingleURLExistsCheck	/DIF/Related_URL/URL		SingleURLExistsCheck	ERROR	Potential Broken Link in ${xpath}	${xpath} is a valid link
SpatialCoordinatesCheck	/DIF/Spatial_Coverage	east is Easternmost_Longitude | north is Northernmost_Latitude | south is Southernmost_Latitude	Spatial Coordinates Check	ERROR	The coordinates in Spatial Coverage are not valid values	
SuffixCheck	/DIF/Spatial_Coverage /Maximum_Depth	postfixes using [m;km]	Suffix Check	ERROR	Unit of measurement for ${name} should be 'm' or 'km.' Units are case sensitive	An appropriate unit is used for ${name}
UniqueEntryIDCheck	/DIF/Entry_ID		Unique EntryID Check	ERROR	${name} is not unique. If this is a new record, please supply a unique ID	${name} is unique
URLFormatCheck	/DIF/Related_URL/URL		URL Format Check	ERROR	URL is not formatted properly in ${xpath}	URL is formatted properly in ${xpath}