WMO Core Metadata Profile 2, Key Performance Indicators

World Meteorological Organization

Date: 2023-10-20

Version: 2.0.0-DRAFT-2023-10-20

Document location: https://community.wmo.int/wis-metadata-kpis

Task Team on WIS Metadata (TT-WISMD)[1]

Expert Team on Metadata Standards (ET-Metadata)[2]

Standing Committee on Information Management and Technology (SC-IMT)[3]

Commission for Observation, Infrastructure and Information Systems (INFCOM)[4]

Table of Contents

Chapter 1. Overview

1.1. Purpose

This document is intended to define Key Performance Indicators (KPIs) in support of the WMO Core Metadata Profile (WCMP). KPIs provide measurable and valuable quality assessment rules over and above the rulesets put forth by WCMP.

The core driver of WCMP KPIs is continuous improvement and useability of discovery metadata as part of the WMO Information System (WIS).[5]

1.2. Scope

This document is bound to the WCMP 2 specification and codelists. All other metadata specifications or representations are not in scope.

1.3. Audience

The target stakeholder audiences for this document include (but are not limited to):

- Metadata providers (NCs, DCPCs)
- WIS2 Global Discovery Catalogues (GDCs)
- WIS2 Nodes
- GAW World Data Centres (WDCs)
- WIS2 Monitoring
- Metadata implementors (generation, ingest)

1.4. How to use

The KPIs in this document are designed to help metadata providers in the curation of discovery metadata, as well as GDCs to measure the quality of metadata from data providers.

In order to improve quality:

- providers should use the KPIs to build into their metadata generation
- WIS2 Global Services and consumers should use the KPIs in order to quality assess discovery metadata and provide subsequent feedback to providers

1.5. Scoring

Each KPI assesses a number of criteria asssociated with metadata quality, resulting in a raw score, as well as a percentage. This approach supports weighted rubric scoring.

1.6. Reference implementation

The TT-WISMD maintains pywcmp[6], as the reference WCMP validation utility which includes:

- validation against WMO Core Metadata Profile 2, Annex A: Conformance Class Abstract Test Suite (Normative)
- validation against the KPIs described in this document

Documentation on installation, configuration and usage can be found on the pywcmp website.

pywcmp is provided as a resource to the community, under continuous improvement. Contributions are welcome and can be facilitated by the WMO Task Team on WIS Metadata.

1.7. Codelists rules

TODO: pending https://github.com/wmo-im/wcmp2-codelists

1.8. Conventions

1.8.1. Symbols and abbreviated terms

Table 1. Symbols and abbreviated terms

Abbreviation	Term
AJAX	Asynchronous JavaScript and XML
CSV	Comma-separated values
DCPC	Data Collection and Production Centres
DOI	Digital Object Identifier
GAW	Global Atmospheric Watch
GDC	Global Discovery Catalogue
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
INSPIRE	Infrastructure for Spatial Information in the European Community
JSON	JavaScript Object Notation
MIME	Multipurpose Internet Mail Extensions
NC	National Centre
OGC	Open Geospatial Consortium
pywcmp	WMO implementation of WCMP validation
URL	Uniform Resource Locator

Abbreviation	Term
WCMP	WMO Core Metadata Profile
WDC	World Data Centre
WIS	WMO Information System
WMO	World Meteorological Organization
XHR	XMLHttpRequest

[1]   https://community.wmo.int/governance/commission-membership/commission-observation-infrastructures-and-information-systems-infcom/commission-infrastructure-officers/infcom-management-group/standing-committee-information-management-and-technology-sc-imt/expert-team-metadata-0

[2]   https://community.wmo.int/governance/commission-membership/commission-observation-infrastructures-and-information-systems-infcom/commission-infrastructure-national-representatives/infcom-management-group/standing-committee-information-management-and-technology-sc-imt/et-metadata

[3]   https://community.wmo.int/governance/commission-membership/commission-observation-infrastructures-and-information-systems-infcom/commission-infrastructure-officers/infcom-management-group/standing-committee-information-management-and-technology-sc-imt

[4] https://community.wmo.int/governance/commission-membership/infcom

[5] https://community.wmo.int/activity-areas/wmo-information-system-wis

[6] https://github.com/wmo-im/pywcmp

Chapter 2. Good quality title

2.1. WCMP element(s)

- `properties.title`

2.2. What is being measured

The title shall always be populated and follow the principles of WCMP guidance. The length shall not be too short or too long, contain less than 3 acronyms and represented in title case. Spelling and grammar are correct.

2.3. Rationale for measurement

The title is the first element of metadata information displayed and helps with initial identification. Meaningful and relevant information makes it easier for users to understand the resource.

In the context of WIS2 Global Discovery Catalogues, the product title and description are the two most relevant elements in the WCMP metadata record. These two elements are presented to the users in search results as well as the product description page, and need to focus on highlighting the product's key characteristics to assist users with relevant product search results.

2.4. Rules for implementation

Table 2. Good quality title implementation rules

Rule	Score
The title is present	1
The title has 3 words or more	1
The title has 150 characters or less	1
The title has only printable characters (numbers and letters)	1
The words in title are represented in "Title Case"	1
The title contains less than 3 acronyms (words with all upper case)	1
The title *does not* contain bulletin header (regular expression: `[A-Z]{4}\d{2}[\s_]*[A-Z]{4}`)	1
The title passes a basic spellcheck	1

Total possible score: 8 (100%)

2.5. Guidance to score well on this assessment

The title should be as specific as possible. For example, if the product only contains one parameter, this can be stated in the title; however, if the product contains numerous parameters, then a more general term should be used in the title, and the parameters stated elsewhere in the metadata record (description, themes, keywords, etc.).

Chapter 3. Good quality description

3.1. WCMP element(s)

- `properties.description`

3.2. What is being measured

The description shall not be too short or too long and contain no HTML markup. Spelling and grammar are correct. Bulletin templates should not be used to populate the description.

3.3. Rationale for measurement

The description faciliates ease of understanding and discovery and so is a key element of metadata information displayed as part of search results. Extensive and meaningful descriptive information allows for users to both understand and properly evaluate a metadata record and its respective resource in support of data access, visualization and exploitation.

In the context of WIS2 Global Discovery Catalogues, the product title and description are the two most relevant elements in the WCMP metadata record.

3.4. Rules for implementation

Table 3. Good quality description implementation rules

Rule	Score
The description has between 16 and 2048 characters	1
The description contains no markup (HTML)	1
The description passes a basic spellcheck	1
The description *does not* contain bulletin template	1

Total possible score: 4 (100%)

3.5. Guidance to score well on this assessment

The description should provide a clear and concise statement that enables the reader to understand the content of the dataset. For guidance when completing the description, consider the following recommendations:

- State the 'things' that are recorded
- State the key aspects recorded about these things
- State what form the data takes

- State any other limiting information, such as time period of validity of the data

- Add purpose of data resource where relevant (e.g. for survey data)

- Aim to be understood by non-experts

- Do not include general background information

- Avoid jargon and unexplained abbreviations

3.5.1. Relevant recommendations

- Avoid adding a scientific description

- Limit information in the description to the specific resource that is being described

- Describe the contents of the resource and the key aspects and/or attributes that are represented

- Explain briefly what is unique about this resource and, if appropriate, how it differs from similar resources

- Avoid citing external sources to this resource

- Avoid spelling out commonly used acronym which are already understood by the general public

- Spell out uncommon acronyms only once

- Avoid including HTML/CSV tables, extra spaces or other markup to control display of text. Use simple paragraph(s) only

- Avoid copying text from a journal article verbatim. This can lead to copyright violation concerns. Additionally, abstracts for journal articles are not intended to describe the provided resource and do not meet the metadata requirements. Related papers can be referenced from and/or tied to the metadata

- Avoid using future verb tense when possible. Write using present or past tenses

3.5.2. Spell checking recommendations

- Dictionary by Merriam-Webster: America's most-trusted online dictionary[1]

- Cambridge Dictionary | English Dictionary, Translations & Thesaurus[2]

3.6. Examples

References of good description examples can be found in TODO.

---

[1] https://www.merriam-webster.com

[2] https://dictionary.cambridge.org

Chapter 4. Time Intervals

4.1. WCMP Element(s)

- `time.interval`
- (`additionalElements.temporal.interval`)

4.2. What is being measured

Whether a time interval is present and contains an interval with a corresponding resolution.

4.3. Rationale for measurement

Temporal information is a significant characteristic of weather/climate/water data and as such is critical for users to know which period(s) of time is/are covered by products and how often new products are received.

4.4. Rules for implementation

Table 4. Temporal information implementation rules

Rule	Score
The begin is less than or equal to the end or open.	1
Only one of the interval extents may be open (begin or end but not both).	1
For every interval there is a corresponding resolution present.	1

Total possible score: 3 (100%)

4.5. Guidance to score well on this assessment

Ensure that the temporal extent resolution is present in the metadata record.

4.6. Examples

```
"time": {
  "interval" : ["2020-10-30", ".."],
  "resolution": "P1D"
}
```

Chapter 5. Graphic overview for metadata records

5.1. WCMP element(s)

- `$.links[?(@.rel=="preview")]`

5.2. What is being measured

The presence of a `preview` link is checked that it contains a URL to a common web image file type.[1]

5.3. Rationale for measurement

Product graphic overviews provide the user with a high level preview of the product which can assist in a high level assessment and/or evaluation as part of search results presentation.

5.4. Rules for implementation

Table 5. Graphic overview for metadata records implementation rules

Rule	Score
A graphic overview element is present	1
A graphic overview URL resolves successfully	1
A graphic overview URL content is a common web image file type (check MIME type, content header/magic number)	1

Total possible score: 3 (100%)

5.5. Guidance to score well on this assessment

In addition to the presence of the graphic overview image it would also be valuable to provide consistent image dimensions (e.g. 800x800 pixels) such that all images are normalized and scaling/alignment of overivew images can be applied consistently by web applications rendering search results.

Examples of catalogues using such information are here:

- GISC DWD

- EUMETSAT Product Navigator

5.6. Examples

```
{
  "rel": "preview",
  "type": "image/png",
  "title": "Browse graphic",
  "href": "https://example.org/path/to/browse.png"
}
```

[1] https://developer.mozilla.org/en-US/docs/Web/Media/Formats/Image_types#Common_image_file_types

Chapter 6. Links health

6.1. WCMP element(s)

Any property with linked information (URLs).

- `links[*].href`

- `properties[*].href`

- `properties[*].url.href`

6.2. What is being measured

The number of broken links in each individual metadata record. Broken links include links which, when accessed, result in a 4xx or 5xx HTTP error.[1]

Also being measured is the use of HTTPS (with a valid SSL certificate) as the link protocol throughout WIS Metadata.

6.3. Rationale for measurement

Broken links damage the user experience and gives the impression to users that a website is not maintained (88% of the online consumers are less likely to return to a site after a bad experience[2]). In addition, having numerous broken links affects the reputation and rank of your website when indexed by mass market search engines.

HTTPS is increasingly becoming a requirement for numerous agencies as well as the suggested protocol vs. HTTP. Having non-HTTPS links in a WCMP document often leads to mixed content errors in web applications deployed via HTTPS for example, and using AJAX/XHR design patterns. HTTPS supports secure, authoritative and trustworthy links as part of WIS Metadata.

6.4. Rules for implementation

Table 6. Links health implementation rules

Rule	Score
Each valid link present	1
Each valid HTTPS link present	1

Total possible score: (valid links + valid HTTPS links) / (total links * 2) (100%)

6.5. Guidance to score well on this assessment

Ensure that all links resolve and are accessible via HTTPS.

6.6. Examples

```
"links": [
  ...
  {
    "rel": "search",
    "type": "text/html",
    "title": "WOUDC - Data - Station List",
    "href": "https://woudc.org/data/stations"
  },
  ...
]
```

[1] https://httpstatuses.com

[2] https://review42.com/web-design-statistics

Chapter 7. Distribution providers

Metadata records should contain information regarding the distribution provider and how to contact them.

7.1. WCMP element(s)

- `$.properties.providers[?(@.role=="distributor")]`

- `$.properties.providers[?(@.role=="distributor")].name`

- `$.properties.providers[?(@.role=="distributor")].contactInfo.email`

7.2. What is being measured

The presence of distribution provider information and supporting elements.

- distributor contact information (organization, email)

7.3. Rationale for measurement

Information of the distributing provider allows the user to contact the distributor in case of anything related to accessing the data.

7.4. Rules for implementation

By detecting the presence of the distributor provider. See WCMP, clause 7.

Table 7. Distributor information implementation rules

Rule	Score
The distributor provider is included.	1
The distributor contact organization is included.	1
The distributor contact email is included.	1

Total possible score: 3 (100%)

7.5. Guidance to score well on this assessment

- Specify a distributor provider. Note that a distributor provider does not have to be the same as the main point of contact, principal investigator

- Specify an email for the distributor

7.6. Examples

```
"properties": {
  ...
  "providers": [{
    ...
    {
        "name": "WMO Lead Centre for Long-Range Forecast Multi-Model Ensemble",
        "contactInfo": {
            "phone": {
                "office": "+82-2-2181-0486",
                "fax": "+82-2-2181-0489"
            },
            "email": {
                "work": "lc_lrfmme@korea.kr"
            },
            "address": [
                {
                "deliveryPoint": "61 16-GIL YEOUIDAEBANG-RO DONGJAK-GU SEOUL",
                "city": "SEOUL",
                "postal-code": "07062",
                "country": "Republic of Korea"
                }
            ],
            "url": {
                "href": "https://www.wmolc.org/"
            }
        },
        "roles": [
            {
                "name": "distributor"
            }
        ]
    }]
  ...
}
```

Chapter 8. Persistent identifiers

8.1. WCMP element(s)

- `properties.externalIds`

8.2. What is being measured

Whether persistent identifier information is available, can be successfully identified, and provides citation instructions.

8.3. Rationale for measurement

Persistent identifiers allow data to be accessible and citable. They make research data easier to access, reuse and verify, thereby making it easier to build on previous work, conduct new research and avoid duplicating already existing work.

8.4. Rules for implementation

Table 8. Persistent identifiers implementation rules

Rule	Score
The external identifiers object is present	1
At least one scheme is equal to `doi`, `ark`, or `hdl`	1
At least one citation exists as a link object with `rel="cite-as"`	1

Total possible score: 3 (100%)

8.5. Guidance to score well on this assessment

- Provide a persistent identifier
- Provide a 'Cite as' template as a link object with `rel="cite-as"`

8.6. Examples

```
"properties": {
  ...
  "externalIds": [{
    "scheme": "doi",
    "value": "https://dx.doi.org/10.14287/10000001"
  }]
  ...
```

```
}
```

For a citation:

```
"properties": {
    ...
    "links": [
    ...
    {
        "rel": "cite-as",
        "title": "Cite as: WMO/GAW Ozone Monitoring Community, World Meteorological
Organization-Global Atmosphere Watch Program (WMO-GAW)/World Ozone and Ultraviolet
Radiation Data Centre (WOUDC) [Data]. Retrieved [YYYY-MM-DD], from https://woudc.org.
A list of all contributors is available on the website. doi:10.14287/10000004",
        "type": "text/html",
        "href": "https://dx.doi.org/10.14287/10000004"
    }]
  ...
]
```