WMCP Metadata Quality KPIs

This document intend to summarize the different Key performance indicators that have been expressed in the discussion available here: <https://github.com/wmo-im/wmcp/issues/2>

### Compliance to WMO Core Profile and its mandatory elements

They should be derived from WMCP documentation part 2 and the rubrik checker.

To be done

### Good quality title

Titles should be consistent and contain meaningful and useful (from search/discovery perspective) content.

It is difficult to create some automated software assessing the good quality of the title.

May some elements can be verified (avoid acronyms, ….) but the general consensus is that a guidance should be developed for WMCP records.

There are two sources:

Below is the **NOAA Title Guidance:**

Style Guidelines

A good archival title summarizes what, where and when data were collected so that a data consumer can read the title and make an informed decision about learning more about the data.

* Capitalize all major words (nouns, pronouns, verbs, adjectives, adverbs, and some conjunctions).
* Avoid using acronyms unless defined in the abstract, especially for projects or organizations. Exception applies to “NOAA” as our audience is familiar with our acronym.
* If you include an acronym, spell out the meaning of the acronym then put the acronym in parentheses after the meaning. Do not assume that everyone reading the title will know or understand all acronyms or jargon. Some common acronyms may be widely understood within a specific domain.
* Always, spell out the title acronym at least once in the abstract text.

WIS Guidance to be referenced here:

### Good quality abstract

Same as above here, a guidance documentation should be provided and established.

What is the existing guidance source: WMO Guidance to be added

Well know bad templates for abstract should be flagged to be gradually replaced.

For instance the automatically generated bulletin records should be replaced by records having a valid description. This can be easily detected automatically and referenced.

## **EDX01 BUFR bulletin available from EUMP (EUMETSAT - DARMSTADT)**

Satellite Centre EUMETSAT Darmstadt - EUMETSAT

---- The bulletin is coded as BUFR code form: FM 94 (BUFR, Binary universal form for the >representation of meteorological data) . (Refer to WMO No.306 - Manual on Codes for the >definition of WMO international codes) ---- The IEDX01 TTAAii Data Designators decode (2) as: T1 >(I): Observational data (Binary coded) - BUFR. A2 (X): Global Area (area not definable). (2: Refer to >WMO No.386 - Manual on the GTS - Attachment II.5) ---- WMO No.9 - Volume C1 'Remarks' field: >IASI SOUNDING (TEMP AND HUMIDITY) GLOBAL DATA`

### Valid Station Identifier

A station id when provided should a WIGOS identifier long version, ie: <https://oscar.wmo.int/surface//index.html#/search/station/stationReportDetails/0-20000-0-10505>

Is it possible to define when it is applicable? The best would be to detect when a station identifier is missing? Any idea to it ?

When a station identifier is provided then it should be a VALID WIGOS identifier url. This can be verified

### Bulletin metadata availability

It must contain some information regarding its availability (available for 24 hours only as an example).

Where to put that information. Should it be repeated in the abstract ?

### Link for WMOEssential data

All WMOEssential data should have a link. Can it be an explicit link (without login).

Not all the time because some datasets are freely available but require login.

### Bulletins have an ongoing temporal extent

Bulletins should have an ongoing temporal extent ie, from 2019-01-01 til now.

It can be enforced automatically.

### Data Parameter Keyword

A Data Parameter keyword should be defined and systematically added in the metadata.

Do we need to extend the WMCP standard.

Can we use the Data Parameter from http://codes.wmo.int ?

### GraphicOverview for non bulletins metadata records

Check the presence of a graphicOverview thumbnail as it improves greatly the search experience when looking at search results.

### Deadlinks

Detection of dealinks contributing to the overall quality score

### Data Policy Possible combination

Check that the data policy information is following the guidance. A set of available combination is possible and should be enforced.