# Metadata guidance and tools for WIS

## Objective

The WIS catalogue stores metadata from a wide variety of contributors (NC, DCPC…). Even if all metadata have to follow the agreed WMO Core Metadata Profile version, the metadata content and quality significantly vary between producers even for similar type of product/bulletin.

The purpose of this paper is to propose a set of “metadata good practices” in addition of the official WMO Core Profile. It would be necessary to complement this set by further recommendations from IPET-MDRD..

Motivations:

1. Assist the NC and DCPC in metadata creation i.e. what kind of textual description shall be put in different free-text sections, semantic to use, preciseness of bounding box, keywords for recommendation different kinds of bulletins …
2. Harmonize the content and semantic of metadata will allow GISC to improve the semantic search engine on GISC portal.

## Example of heterogeneous metadata

An illustration for land SYNOP and current lack of harmonization:

### Stations listing descriptive keywords

If we focus on the keyword part for stations listing:

|  |  |  |
| --- | --- | --- |
| **AHL** | **Keywords id** | **Keywords extract** |
| SMDL01EDZW | <gmd:descriptiveKeywordsuuidref="place"> | <gco:CharacterString>Nürnberg</gco:CharacterString><gco:CharacterString>10763</gco:CharacterString> |
| SMFR01LFPW | <gmd:MD\_Keywords id="WMONo9VolAKeywords"> | <gco:CharacterString>07005</gco:CharacterString> |
| SMJP01RJTD | <gmd:MD\_Keywords id="placecode"> | <gco:CharacterString>WAKKANAI</gco:CharacterString> |
| ISML01RJTD | <gmd:MD\_Keywords id="placecode-VOLA"> | <gco:CharacterString>89532 SYOWA</gco:CharacterString> |

We can notice that for a same data type (land synop) the free section keyword for station listing is different. Depending of metadata origin the keyword id differs: “place”, “placecode”, “WMONo9VolAKeywords” and it is the same for the content where the station place is sometime the station index or only the station name, or both in same string, or both in different keywords.

In this example we can note that the different way of representation will complicate the efficiency of search engines.

### Synop dissemination descriptive keywords

The same observation can be done for dissemination time, where we can find a “temporal“ descriptive keyword with values: 00 12 18… for others bulletins it will be “updtime-rt” with value : Dissemination 00Z Dissemination 12Z Dissemination 18Z, for others bulletins no information regarding the dissemination …

|  |  |  |
| --- | --- | --- |
| **AHL** | **Keywords id** | **Keywords extract** |
| SMDL01EDZW | <gmd:descriptiveKeywordsuuidref="temporal"> | <gco:CharacterString>18</gco:CharacterString> |
| SMFR01LFPW |  |  |
| ISML01RJTD | <gmd:MD\_Keywords id="updtime-rt"> | <gco:CharacterString>Dissemination 00Z</gco:CharacterString> |

### Metadata SYNOP Title

For title different examples:

1. GTS Bulletin: SMDL01 EDZW - Surface data (details are described in the abstract)
2. Report of surface observation from a fixed land station \_ FRANCE : ABBEVILLE, LILLE-LESQUIN, PTE DE LA HAGUE, CAEN-CARPIQUET, ROUEN-BOOS, 07070, BREST-GUIPAVAS, PLOUMANAC H, RENNES-ST JACQUES, ALENCON, …
3. WIS/GTS bulletin ISML01 RJTD in FM94 BUFR/Edition 4 (SYNOP)
4. SMBZ01 SBBR main synoptic hour surface observations
5. SMBE01 collection of SYNOP reports available from TXKF (BERMUDAINTERNATIONALAIRPORT) at 00, 06, 12 and 18 UTC

## Proposition for metadata harmonization

**Metadata template by data type**

A template describing the expected content for each metadata field depending on data type would vastly improve the harmonization of metadata and consequently the efficiency of the search on GISCs portals.

Such a work has been started by IPET-MDRD (<http://www.wmo.int/pages/prog/www/WIS/wiswiki/tiki-index.php?page=IndexForMetadataGuidance&pagenum=4#Example_metadata_records>) and a template example for SYNOP land surface has been provided: <http://www.wmo.int/pages/prog/www/WIS/wiswiki/tiki-index.php?page=MDG_Land>

### Example of guidance provide by IPET-MDRD:

1. **for bounding box field**: *Ranges of latitude and longitude must be indicated. That is WIS requirement. Please note that "done is better than perfect". The bounding box is useful because any kinds of geospatial data have in common. So impreciseness is built-in virtue, not problem. Don't worry about preciseness - it has no use to spend too much time about 0.01 degree digit, discussing geodetic datums, or dreaming about polygon or multipoint extensions. Even if you don't know exact city of the observing station or if it is moving frequently, that's no problem. Simply fill the bounding box of rough maximum possible range of lat/longs.*
2. **for title** : *The* gmd:title *element is the title of the dataset by which the dataset is cited. For GTS bulletins, in addition to the general guideline given in* [MDG\_DO](http://www.wmo.int/pages/prog/www/WIS/wiswiki/tiki-index.php?page=MDG_DO)*, it is recommended to include the heading TTAAii CCCC.*

This guidance is a very good starting point for the template, but the content for some elements deserves being further improved. For example, on the “title” element, the template should specify the semantic to use (synop or synoptic or surface data), recommendation on the language to use, recommended the string order (example first what then where and when…)...

The purpose of the template is neither to impose a semantic and a format nor to modify all the metadata that have been previously created but to provide recommendations and templates on which metadata editor can build new metadata.

### Metadata validation tool

At present time the way for metadata validation is different from each GISC, as well as the granularity checking. It would be useful to have an official common “WMO Core schematron” that should be used (and improved) by all the GISC to check the quality of the metadata structure before insertion on catalogue. By sharing a common schematron we would be sure that any metadata insert on any GISC set would be correct.

### Thesaurus on data type designator

To facilitate advanced search, a thesaurus based on “WMO ATTACHMENT II–5” data type keywords, at least T1T2(A1) header shall be proposed in order for metadata editor to have the possibility to associate a metadata data to a specific data type , sample of useful thesaurus keywords:

“SYNOP“, “SYNOP MOBIL”, “CLIMAT TEMP SHIP”, “TEMP”, “GRIB”…

## References

## Recommended Text For Report

*To be discussed and reformulate if necessary:*

TT-GISC noted the recommendation that the metadata content should be harmonized in order to improve the semantic search on GISC portal. TT-GISC expresses the need of metadata template by data type to assist user in metadata creation and guarantee the harmonization of content.

In addition to encouraging GISCs to apply a minimum common use of keywords in guiding its centres in creation of metadata, the WIS guidelines are too broad to support such as practice. Furthermore, TT-GISCs would like to see an optimization of metadata still providing the necessary and sufficient information that will describe the data or product.

TT-GISC noted the informal work that has been started on templates done by IPET-MDRD. TT-GISC encourages to IPET-MDRD to continue this work and deliver a first version to TT-GISC as soon as possible.

TT-GISC noted that these templates would be recommendations on how metadata shall be written.

TT-GISC ~~expressed~~ acknowledged the need of a common “WMO Core schematron” for metadata validation before insertion on GISC catalogue. This schematron would be shared and improved by the GISC community.

TT-GISC highlighting that we have concerns about the different result lists that are caused by lack of harmonizating the application of keywords, and requests IPET-MDRD to take into consideration the recommendations of this paper provided to TT-GISC and to consider provision of ~~provide~~ a thesaurus on data type designator based on the keywords from “WMO ATTACHMENT II–5”. The purpose of this thesaurus is to classify metadata by data/product type and facilitate search for meteorological experts.

--------------------

# Annex

## Synop metadata example from IPET-MDRD wiki

Please note following record is made for example purpose and operational record may be different.

**MD\_Metadata**
**fileIdentifier**: urn:x-wmo:md:int.wmo.wis::ISMC01RJTD
**language**: eng
**characterSet**: utf8
**hierarchyLevel**: dataset
**contact**:
  **CI\_ResponsibleParty**
  **@id**: wis\_centre
  **organisationName**: WIS/NC Japan - Japan Meteorological Agency
  **contactInfo**:
    **CI\_Contact**
    **phone**:
      **CI\_Telephone**
      **voice**: tel:+81-3-3212-8341
    **address**:
      **CI\_Address**
      **electronicMailAddress**: wis-jma@met.kishou.go.jp
  **role**: pointOfContact
**dateStamp**: 2014-11-18T00:00:00Z
**metadataStandardName**: WMO Core Metadata Profile of ISO 19115 (WMO Core), 2003/Cor.1:2006 (ISO 19115), 2007 (ISO/TS 19139)
**metadataStandardVersion**: 1.3
**identificationInfo**:
  **MD\_DataIdentification**
  **citation**:
    **CI\_Citation**
    **title**: ISMC01 RJTD - Surface Synoptic Reports on main synoptic hours in Japan
    **date**:
      **@nilReason**: inapplicable
    **identifier**:
      **MD\_Identifier**
      **code**: urn:x-wmo:md:int.wmo.wis::ISMC01RJTD
  **abstract**: Surface synoptic observation on main synoptic hours (00, 06, 12, and 18 UTC) from fixed land stations in Japan, as defined in RBSN (Regional Basic Synoptic Network). Place: 47401 (WAKKANAI) 47412 (SAPPORO) 47420 (NEMURO) 47426 (URAKAWA) 47582 (AKITA) 47590 (SENDAI) 47600 (WAJIMA) 47648 (CHOSHI) 47662 (TOKYO) 47678 (HACHIJOJIMA) 47740 (SAIGO) 47778 (SHIONOMISAKI) 47807 (FUKUOKA) 47827 (KAGOSHIMA) 47843 (FUKUE) 47893 (KOCHI) 47909 (NAZE) 47918 (ISHIGAKIJIMA) 47936 (NAHA) 47945 (MINAMIDAITOJIMA) 47971 (CHICHIJIMA) 47991 (MINAMIORISHIMA). The dataset is provided in FM 94 BUFR.
  **pointOfContact**:
    **CI\_ResponsibleParty**
    **@id**: content\_contact
    **organisationName**: WIS/NC Japan - Japan Meteorological Agency
    **contactInfo**:
      **CI\_Contact**
      **phone**:
        **CI\_Telephone**
        **voice**: tel:+81-3-3212-8341
      **address**:
        **CI\_Address**
        **electronicMailAddress**: wis-jma@met.kishou.go.jp
    **role**: originator
  **resourceFormat**:
    **MD\_Format**
    **@id**: format.r1
    **name**: FM 94 BUFR
    **version**: XIV
  **descriptiveKeywords**:
    **MD\_Keywords**
    **@id**: subjkey
    **keyword**: synopticMeteorology
    **keyword**: weatherObservations
    **type**: theme
    **thesaurusName**:
      **CI\_Citation**
      **title**: WMO\_CategoryCode
      **date**:
        **@nilReason**: template
  **descriptiveKeywords**:
    **MD\_Keywords**
    **@id**: placecode
    **keyword**: 47401 - WAKKANAI
    **keyword**: 47412 - SAPPORO
    **keyword**: 47420 - NEMURO
    **keyword**: 47426 - URAKAWA
    **type**: place
    **thesaurusName**:
      **CI\_Citation**
      **title**: WMO Publication No. 9, Volume A
      **date**:
        **@nilReason**: unknown
  **descriptiveKeywords**:
    **MD\_Keywords**
    **@id**: updtime
    **keyword**: 00:00:00Z
    **keyword**: 06:00:00Z
    **keyword**: 12:00:00Z
    **keyword**: 18:00:00Z
    **type**: temporal
  **descriptiveKeywords**:
    **MD\_Keywords**
    **@id**: themekey
    **keyword**: pressure
    **keyword**: pressureReducedToMeanSeaLevel
    **keyword**: pressureTendency
    **keyword**: airTemperature
    **keyword**: dewPointTemperature
    **keyword**: horizontalVisibility
    **keyword**: totalPrecipitation
    **keyword**: cloudAmount
    **keyword**: cloudType
    **keyword**: weather
    **keyword**: windDirection
    **keyword**: windSpeed
    **type**: theme
  **resourceConstraints**:
    **MD\_LegalConstraints**
    **accessConstraints**: otherRestrictions
    **useConstraints**: otherRestrictions
    **otherConstraints**: GlobalExchange
    **otherConstraints**: GTSPriority2
    **otherConstraints**: WMOEssential
  **language**: eng
  **topicCategory**:
    **MD\_TopicCategoryCode**
  **extent**:
    **EX\_Extent**
    **@id**: boundingExtent
    **geographicElement**:
      **EX\_GeographicBoundingBox**
      **@id**: boundingGeographicBoundingBox
      **westBoundLongitude**: 124.1667
      **eastBoundLongitude**: 153.9833
      **southBoundLatitude**: 24.2833
      **northBoundLatitude**: 45.4167
**distributionInfo**:
  **MD\_Distribution**
  **distributionFormat**:
    **MD\_Format**
    **@id**: format.1
    **name**: FM 94 BUFR
    **version**: XIV
  **distributor**:
    **MD\_Distributor**
    **distributorContact**:
      **CI\_ResponsibleParty**
      **organisationName**: WIS/GISC Tokyo
      **role**: distributor
    **distributorTransferOptions**:
      **MD\_DigitalTransferOptions**
      **onLine**:
        **CI\_OnlineResource**
        **linkage**: http://www.wis-jma.go.jp/gisccache/urn:x-wmo:md:int.wmo.wis::ISMC01RJTD
        **function**: download