# Discovery Metadata Management

## Introduction

### Function of Discovery Metadata

1. Discovery metadata is
   1. records containing various characteristics associated with a data or product and
   2. basic and core database to implement the Discovery Access and Retrieve (DAR) catalogue, which provides a means for locating and retrieving data and products throughout WIS.
2. The WIS Technical Regulations (WMO-No. 49) paragraph A.3.3 states: *WIS functions and operation shall be based on catalogues that contain metadata describing data and products available across WMO, plus metadata describing dissemination and access options*.
3. Metadata for any data or product available from a DCPC or GISC is to be created or updated in the DAR Catalogue of the DCPC or GISC, before the data or product is available.

### Standards and Specifications Conformance of Discovery Metadata

1. WMO Core Metadata Profile (WCMP) Specification version 1.3, approved by Presidents of Technical Commissions 2013, states:

*Authors of discovery metadata records published within the WIS DAR Catalogue are required to comply with the WMO Core Metadata Profile. Thus WIS Discovery Metadata shall be compliant with:*

* *ISO 19115:2003 ‘Geographic information – Metadata’;*
* *ISO 19115:2003/Cor. 1:2006 ‘Geographic information – Metadata – Corrigendum 1’; and*
* *Additional constraints described in this Manual (WCMP Specification).*

*Specifications in this Manual (WCMP v1.3 Specification) shall take precedence over the specifications in ISO 19115:2003 and ISO 19115:2003/Cor. 1:2006.*

***6.1.1 Each WIS Discovery Metadata record shall validate without error against the XML schemas defined in ISO/TS 19139:2007.***

***6.1.2 Each WIS Discovery Metadata record shall validate without error against the rule-based constraints listed in ISO/TS 19139:2007 Annex A (Table A.1).***

## Discovery Metadata Lifecycle Management

1. In order to provide Discovery, Access and Retrieval services based on metadata, discovery metadata is managed continuously throughout their lifecycle: information extraction, generation, publish, exploitation, quality evaluation, and revision.

Information Extraction

Metadata Generation

Metadata Exploitation

Metadata Revision

Metadata Quality Evaluation

Metadata Publish

Metadata Model

1. Discovery metadata lifecycle management is a continuous process, based on metadata model, by means of: composing, organization, maintenance, and quality assurance.

## Discovery Metadata Model

1. The metadata model defines the specifications and application of the metadata, describing the generic framework of the metadata. The metadata model is created and updated in accordance with metadata specifications, characteristics of described data, guidance of experts, agreed approach of application, and good practices. The metadata model is the framework determining set of metadata.
2. The metadata model is:
   1. sketched up on information extraction from materials, such as WMO No.9 Vol. C1, WMO No.9 Vol. A, and contacts, etc.;
   2. used by metadata generation to produce metadata set;
   3. used by metadata quality evaluation to validate, monitor, evaluate the metadata quality on syntax, schema, semantics, and usability;
   4. updated by metadata revision to promote the latest evaluation outcome.
3. As the generic framework of metadata, metadata model conforms to the following standards, and specifications:
   1. ISO 19115:2003/Cor. 1:2006 ‘Geographic information – Metadata – Corrigendum 1’,
   2. ISO/TS 19139:2007 Annex A,
   3. WMO Core Metadata Profile (WCMP) Specification.
4. As the generic framework of metadata for WMO DAR services, metadata model is inclined to comply with the application guidance.

ISO 19115:2003/Cor. 1:2006

ISO/TS 19139:2007 Annex A

WMO Core Metadata Profile (WCMP) Specification

WMO Application Guidance of Discovery Metadata

Discovery Metadata Model

Discovery Metadata records

1. The Application Guidance of Discovery Metadata will provide guidance on the following topics, each topic corresponding to one discovery metadata model: (TT-ApMD-2 Final Report Annex C)
   1. World Wide Watch
      * Surface synoptic
      * Upper air
      * Satellite radiances
      * Scatterometer
      * Radar altimeter waves
      * SAR waves
      * Satellite water vapour
      * GPS water vapour
      * Weather radar
      * NWP weather forecast
      * Weather alert
      * Weather warning
   2. Aviation
      * TAF
      * METAR
      * SigWx
   3. Climatology
      * CLIMAT
      * Monthly/seasonal/annual/decadal analyses
      * IPCC reports
      * Climate simulations
   4. Oceanography
      * Sub-surface observations (physical)
      * Biological observations
      * Surface observations
      * Chemistry observations
   5. Cryosphere
      * Sea ice
      * Glaciers
      * Snowfall
   6. Atmospheric chemistry
      * Radiative measurements
      * Chemical concentrations
   7. Hydrology
   8. Space Weather

## Discovery Metadata Composing

1. Discovery metadata composing includes two phases: model instantiation and metadata generation.
   1. Metadata model instantiation is the process of specifying the “blank” fields in the model framework, in accordance with the model guidance, and completing a prototype metadata. In this phase, metadata composing focuses on specific one metadata.
   2. Metadata generation is the process of producing the metadata of the same model as the prototype. This process is usually conducted by automatic generation tool. The tool is able to complete almost all fields based on the prototype, and prompt metadata author (human) to revise and confirm the generated metadata. All metadata should be confirmed by metadata author before being published.
   3. The sources for automatic metadata composing are catalogue databases, like WMO No. 9 Vol. C1, Vol. A, etc. Those catalogue databases are digested and key information is extracted and injected into the metadata.
2. The discovery metadata, which is intended to substitute Vol. C1, should contain all necessary information of Vol. C1, and provides additional information for DAR services. When the discovery metadata is used, instead of Vol. C1, for discovery, access and retrieval, they should provide convenient way for DAR to the whole information in Vol. C1.
3. Task Team on Application of Metadata (TT-ApMD) is developing the guidance on writing discovery metadata records providing the information needed to fulfill the use cases form WMO No. 9 Vol. C1. The guidance includes metadata templates, good examples, validation tools, spreadsheets for automatic generation, documents and presentations for users and metadata experts. TT-ApMD will create standard templates for GTS reports. (TT-ApMD-2 Final Report)
4. The discovery metadata composing is an automatic process with expert’s revision, as illustrated below.

Automatic Metadata Generation Tool

Metadata Validation Rules

Metadata Prototype

Metadata Model

Catalogue Database

Metadata Generation

Metadata set



Metadata Expert

*revision*

*revision*

## Discovery Metadata Organization

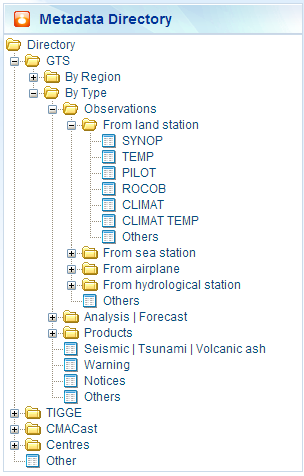
1. Discovery metadata organization provides convenience of metadata management. The main approaches are thesaurus-based, hierarchy-based and collection-based organizations.
2. Thesaurus-based approach employs the *keywords groups* (gmd:MD\_Metadata / gmd:identificationInfo / gmd:MD\_Identification / gmd:descriptiveKeywords element) for organizing the metadata: each keywords group describes with keywords in same thesaurus; different groups describe the data in different thesauri. For example, station identifiers are used as keywords with Vol. A as thesaurus; time group in Vol. C1 are used as keywords with time group thesaurus. In WCMP v1.3, a new code list WMO\_CategoryCode is introduced to keywords, and requires:

***8.2.1 Each WIS Discovery Metadata records shall include at least one keyword from the WMO\_CategoryCode code list.***

***8.2.2 Keywords from WMO\_CategoryCode code list shall be defined as keyword type “theme”.***

***8.2.3 All keywords sourced from a particular keyword thesaurus shall be grouped into a single instance of the MD\_Keywords class.***

1. Hierarchy-based approach exploits *hierarchyLevelName* element to supply the taxonomy information, and populates a hierarchical tree on WIS portal website. The users can find the metadata through browsing, rather than searching. The metadata directory in GISC Beijing provides such hierarchy service, as an example below. IPET-MDRD will consider introducing a mechanism for browsing the catalogue.



1. Collection-based approach attempts to combine the metadata into collections as summary metadata records. This is particularly helpful and convenient to DAR the relative metadata as whole group, like numerical weather prediction products and satellites products. The practice and experience reported by GISC Tokyo shows that the content in collection is determined by user requirements, which is based on user survey.

## Discovery Metadata Maintenance

1. Discovery metadata should be consistent with present situation of data and metadata specifications, and need to be maintained in time. There are generally 7 situations that need metadata maintenance:
   1. data specification updates,
   2. contact information updates,
   3. distribution information updates,
   4. error correction,
   5. data discontinue,
   6. affiliation changes, and
   7. standards and specifications of metadata updates.
2. In the first five cases above, it is the responsibility of metadata experts (point of contact) to update the metadata records in time. The errors that require correction include XML validation fails, editorial mistakes, and other cases that metadata records are in low quality.
3. In the next two cases, it is the responsibility of metadata experts and principle GISCs to revise the metadata; re-generating the metadata records if necessary. This turns to metadata generation.

## Discovery Metadata Quality Assurance

1. The quality assurance of discovery metadata covers both composing phase and ongoing health checks.
   1. In metadata composing phase, metadata validation tool makes sure the generated metadata records conform to the standards and specifications by schematron and test cases. The schematron and “rubric” test developed by NOAA shows a good solution.
   2. The schematron and “rubric” test hosted by NOAA websites can also validate collections of metadata. In the future version, this validation service will be able to harvest metadata and offer continuous health checks. It is recommended to provide metadata health check service.
2. The metadata records are evaluated to be in low quality need be revised and updated.

## References

* [1] The Manual on WIS (WMO No. 1060), http://www.wmo.int/pages/prog/www/WIS/documents/Manual-on-WIS-en.pdf‎
* [2] The Guide to WIS (WMO No. 1061), http://www.wmo.int/pages/prog/www/WIS/documents/Guide-to-WIS-en.pdf‎
* [3] WIS Compliance Specification of GISC, DCPC, and NC, ftp://ftp.wmo.int/Documents/wis/WIS-TechnicalSpecification-v1-0.doc
* [4] Discovery Access and Retrieval (DAR), https://www.wmo.int/pages/prog/www/WIS/DAR\_en.html
* [5] Catalogue of Meteorological Bulletins (WMO No. 9 Vol. C1), https://www.wmo.int/pages/prog/www/ois/Operational\_Information/VolC1\_en.html
* [6] Observing Stations (WMO No. 9 Vol A), http://www.wmo.int/pages/prog/www/ois/volume-a/vola-home.htm
* [7] The Manual on Codes (WMO No. 306), http://www.wmo.int/pages/prog/www/WMOCodes.html

## Recommended Text

Include here suggested text for inclusion in Report, including and proposed amendments to the Manual or Guide to WIS and Manual on GTS.

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