

Climate Data Model

for a Climate data management system

Contents

[A way forward 3](#_Toc56353133)

[Overview / Summary 3](#_Toc56353134)

[Review of existing data models 3](#_Toc56353135)

[Data models considered 3](#_Toc56353136)

[Survey results for hourly observations data models 3](#_Toc56353137)

[Awaiting answer from 4](#_Toc56353138)

[List of full schemas shared on the OpenCDMS GitHub 4](#_Toc56353139)

[Assessment of existing data models 4](#_Toc56353140)

[Structure of the hourly data table 4](#_Toc56353141)

[Data Quality codes 5](#_Toc56353142)

[Assessment on existing Climatological Practices 6](#_Toc56353143)

[Computation practices and rules with missing data 6](#_Toc56353144)

[Main Principles 6](#_Toc56353145)

[Able to manage Essential Climate Variables (ECV) 6](#_Toc56353146)

[Compliant with the GCOS Climate Monitoring Principles 7](#_Toc56353147)

[Compliant with the WIGOS Metadata 7](#_Toc56353148)

[Compliant with the CIMO Guide 7](#_Toc56353149)

[Compliant with the Manual on the HQ GDMM 7](#_Toc56353150)

[Compliant with the CDMS Specifications 8](#_Toc56353151)

[Compliant with current data standards 9](#_Toc56353152)

[WMO 9](#_Toc56353153)

[WIS 9](#_Toc56353154)

[METCE 9](#_Toc56353155)

[WMO Register Codes 9](#_Toc56353156)

[WMO and OGC 9](#_Toc56353157)

[Met Ocean DWG 9](#_Toc56353158)

[Hydrology Domain Working Group 10](#_Toc56353159)

[OGC and ISO 10](#_Toc56353160)

[Recommendations for next generation data models 10](#_Toc56353161)

[Research questions 10](#_Toc56353162)

[Data Model Considerations 10](#_Toc56353163)

[ Primary keys and indexing 10](#_Toc56353164)

[ Date period 10](#_Toc56353165)

[ Long vs wide 10](#_Toc56353166)

[ Dynamical schema modifications 11](#_Toc56353167)

[References 11](#_Toc56353168)

[Climate Data Management System Specifications 11](#_Toc56353169)

[Guide to Instruments and Methods of Observation 11](#_Toc56353170)

# A way forward

# Overview / Summary

Define “climate” (reference GCOS variables)

Mention alternative terminology including “earth systems”

# Review of existing data models

Data models considered

### Survey results for hourly observations data models

| **#** | **Country** | **Name** | **Type** | **System Name** |
| --- | --- | --- | --- | --- |
|  | AUSTRALIA | BoM | NMHS | CLIDE |
|  | BRAZIL | INMET | NMHS | INMET |
|  | CANADA | MSC | NMHS | ESC |
|  | CZECH | CHMI | ATACO | CLIDATA |
|  | DWD | DWD | NMHS |  |
|  | FRANCE | MF | NMHS | BDCLIM |
|  | FRANCE | MFI | METEO FRANCE INTERNATIONAL | CLISYS |
|  | NEW ZEALAND | METSERVICE | NMHS | CLIDB |
|  | RUSSIA | HYDROMET | NMHS | CLIWARE |
|  | SPAIN | AEMET | NMHS | SAEMET |
|  | UK | UK Met Office | ORGANIZATION | CLIMSOFT |
|  | UK | UK Met Office | NMHS | MIDAS |
|  | WMO | WMO | ORGANIZATION | CLICOM |
|  | WMO | WMO | ORGANIZATION | MCH |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

### Awaiting answer from

| **Country** | **Name** | **Type** | **System Name** |
| --- | --- | --- | --- |
| AUSTRALIA | BoM | NMHS | ADAM |
| CHINA | CMA | NMHS |  |
| JAPAN | JMA | NMHS |  |
| SLOVAKIA | Private | IMS | CLDB |

### List of full schemas shared on the OpenCDMS GitHub

|  |  |  |
| --- | --- | --- |
| **Country** | **Name** | **System Name** |
| AUSTRALIA | BoM | CLIDE |
| FRANCE | MF | BDCLIM |
| UK | UK Met Office | CLIMSOFT |
| UK | UK Met Office | MIDAS |
| ORGANIZATION | WMO | CLICOM |
| ORGANIZATIO | WMO | MCH |

# Assessment of existing data models

## Structure of the hourly data table

Following the terms used by WMO-TD No. 60 (2007) where 3 data model types have been presented (Element Model, Observation Model, Value Model)

|  |  |
| --- | --- |
| Element model (EM) | An Element Model represents data in tables, having, in each row, different values of one variable observed at one station at different times.  For example, hourly data could be stored in an **Houly** table. Each row would correspond to a specific station, a specific day, and a specific variable. The attributes, i.e. each cell of a specific row, store the different values of that variable and of that station for a given hour (e.g. 24 values for 1 day). |
| Observation model (OM) | An Observation Model represents data in tables having, in each row, the values of different variables observed at one station at a given time.  For example, hourly data could be stored in an **Hourly** table. Each row would correspond to a specific station at a specific hour. Each column of a specific row would store the values of the different hourly variables observed at the specific hour, e.g. hourly max temperature, hourly mean temperature and precipitation. |
| Value model (VM) | A Value Model will represent the data values in tables having, in each row, only one value of one variable observed at one station at a specific time.  For example, hourly data could be stored in an Hourly. Each row would correspond to a specific station at a specific time, for a specific variable and contains only 1 value |

| **Country** | **Data Model Name** | **Data Model Type** |
| --- | --- | --- |
| AUSTRALIA | CLIDE | OM |
| BRAZIL | INMET | ? |
| CANADA | ESC | VM |
| CZECH | CLIDATA | EM |
| DWD |  | VM |
| FRANCE | BDCLIM | OM |
| FRANCE | CLISYS | OM |
| NEW ZEALAND | CLIDB |  |
| RUSSIA | CLIWARE | OM |
| SPAIN | SAEMET | EM |
| UK | CLIMSOFT | VM |
| UK | MIDAS | OM |
| WMO | CLICOM | EM |
| WMO | MCH | VM |

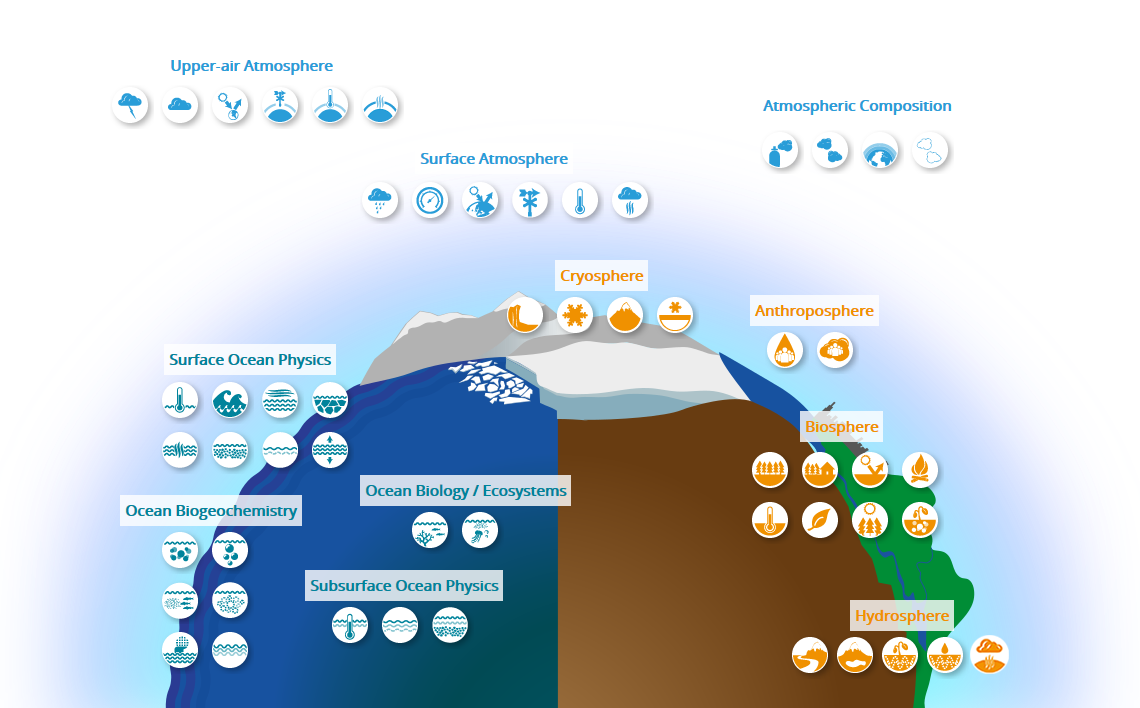
## Data Quality codes

# Assessment on existing Climatological Practices

## Computation practices and rules with missing data

# Main Principles

## Able to manage Essential Climate Variables (ECV)





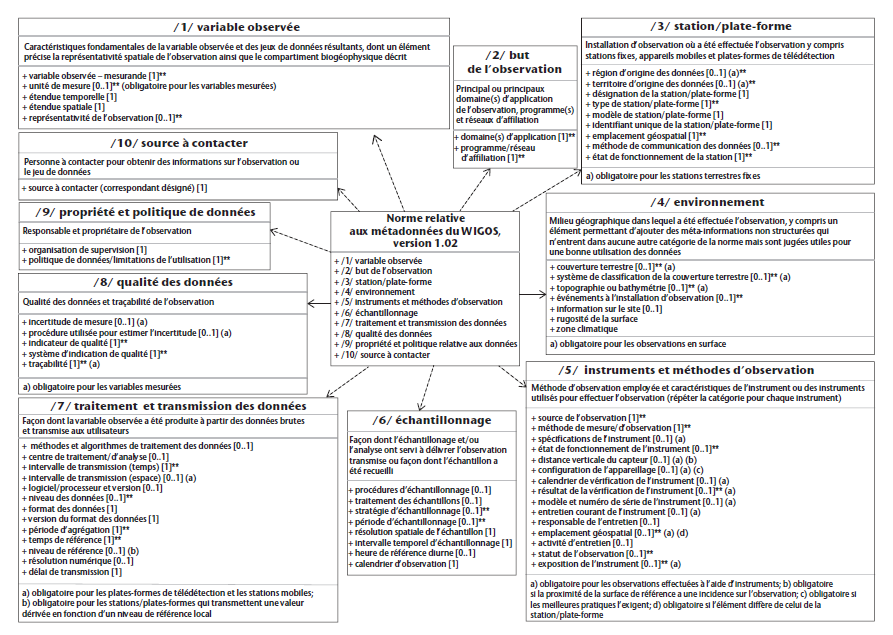
<https://gcos.wmo.int/en/essential-climate-variables/table>

Inventory of ECV Inventory : <https://climatemonitoring.info/wp-content/uploads/2020/07/ECV_Inventory_v3.0.xlsx>

## Compliant with the GCOS Climate Monitoring Principles

See : <http://ane4bf-datap1.s3-eu-west-1.amazonaws.com/wmocms/s3fs-public/ckeditor/files/GCOS_Climate_Monitoring_Principles.pdf?l1e4ALNYxVIStmm19we2Sz0evxEFpHmT>

## Compliant with the WIGOS Metadata



**See** WMO- No. 1192

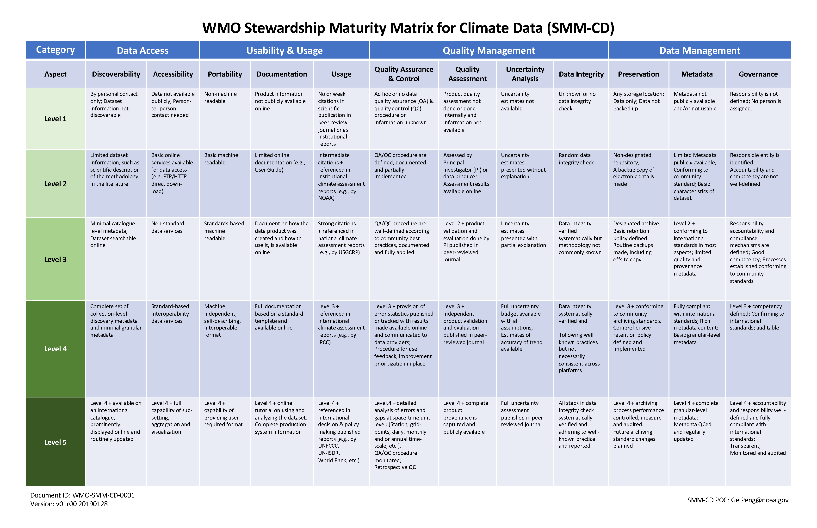
## Compliant with the CIMO Guide

Guide to Instruments and Methods of Observation (WMO- No. 8°)

And especially the **Sitting Classification** (first common ISO/WMO standard published by ISO as ISO standard 19289:2014 (EN)).and the **Sustained Performance maintenance**

## Compliant with the Manual on the HQ GDMM

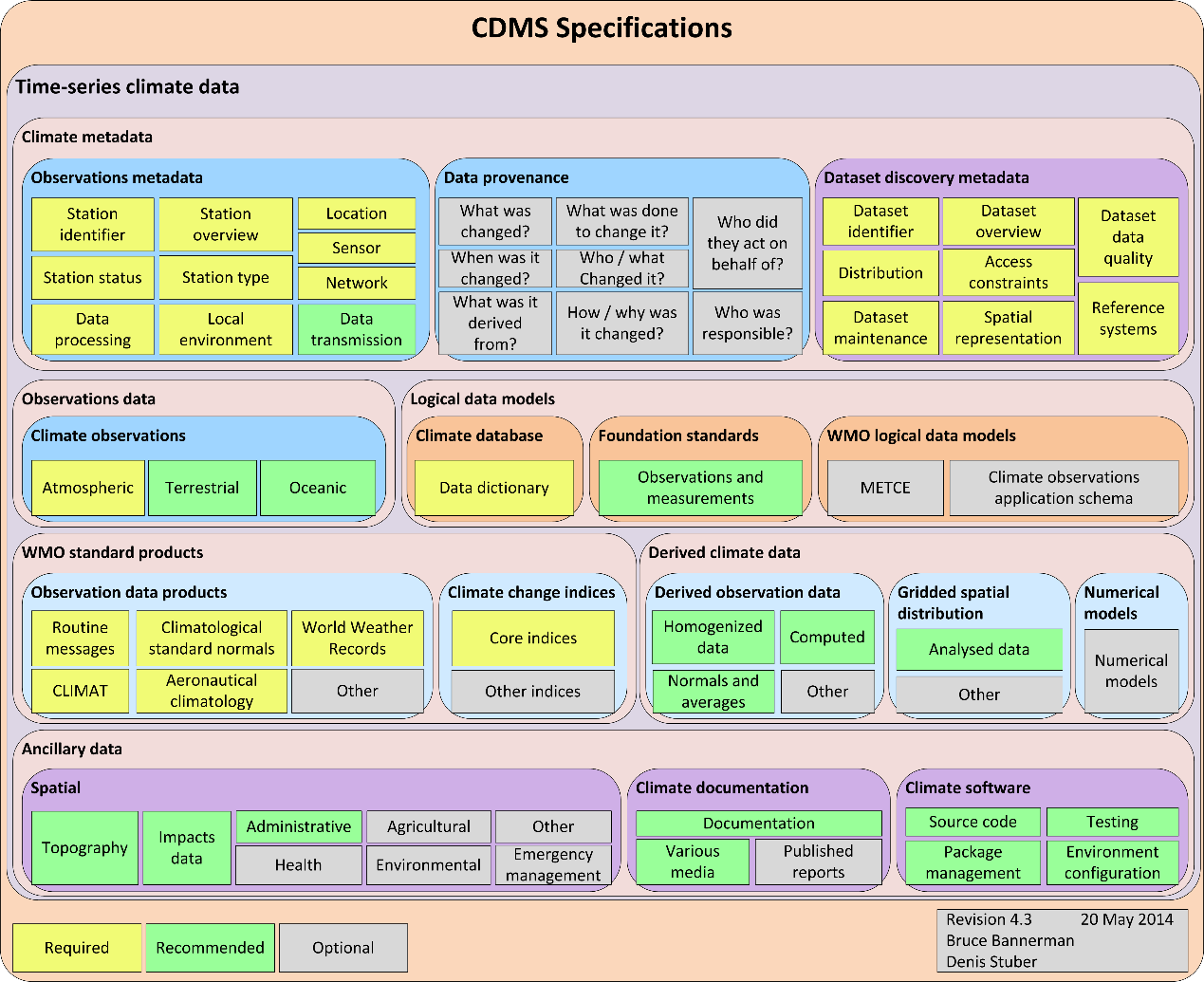
High-quality Global Data Management Framework for Climate WMO- No. 1238. Especially with the International and National WMO Stewardship Maturity Matrix for Climate Data.



## Compliant with the CDMS Specifications

Climate Data Management System Specifications WMO- No. 1131

See mainly the chapter 4 of the CDMS Specifications



* Handling observations from multiple sensors per station, per phenomenon, and recording the source of each observation.
* Managing multiple tiers of data quality, from raw records to homogenized data.
* **Managing spatial and time-series data**
* Using a robust data model that takes into account the requirements of open spatial standards, particularly the ISO 19156:2011 *Geographic information – Observations and measurements* standard, METCE and the WMO climate observations application schema (see component 4.2.3.2).
* Managing metadata related to data provenance. This entails ensuring that each change to an observation is recorded for future recovery, and recording the details of why a particular change was made, which includes:
  + Tracing the product lineage to the data source. For example, what observations and gridded data were used to underpin the analysis released in peer-reviewed paper X?
  + Ensuring that the reason for each observation change is recorded.
* Managing third-party and crowdsourced data.
* Managing intellectual property rights related to data.
* Enabling point-in-time recovery. For example, what data were present in the database for station X at time T?
* Storing a range of document formats, such as:
  + Photographs of observation stations and instruments, meteorological phenomena, etc.
  + Scanned paper observation forms
  + Scanned microfiche/microfilm
  + Relevant observations metadata documents, such as instrument calibration reports
  + Technical manuals
  + Site location plans and sections
  + Videos and other multimedia formats
* Handling data uncertainty (for more information, see Wikipedia articles on uncertain data and uncertainty).
* Managing multidimensional time-series gridded data and possibly numerical models.
* Providing support for the information management concepts of semantics and linked data.

## Compliant with current data standards

Not only station time series but able to manage **spatial** data. More and more data are created for climate services and should be available for NMHSs : models data, reanalysis, radars, etc.

## WMO

### WIS

### METCE

### WMO Register Codes

## WMO and OGC

### Met Ocean DWG

WMO and Open Geospatial Consortium (OGC): The Meteorology and Oceanography Domain Working Group (Met Ocean DWG)

### Hydrology Domain Working Group

[WMO et Open Geospatial Consortium (OGC): The Hydrology Domain Working Group](http://external.opengis.org/twiki_public/HydrologyDWG/WebHome)

## OGC and ISO

**ISO 19156:2011** *Geographic information – Observations and measurements. E.g.* WaterML and the Meteorological Information Exchange Model of the International Civil Aviation Organization (ICAO).

[**ISO 19115**](https://www.iso.org/fr/standard/67039.html) *Geographic Information - Metadata standard* (not for station metadata)

[**ISO 19131**](https://www.iso.org/fr/standard/36760.html) *Geographic information – Data product specifications*

[**ISO 19157**](https://www.iso.org/fr/standard/32575.html) *Geographic Information - Data Quality*

[**ISO 19158**](https://www.iso.org/standard/32576.html) *Geographic Information - Quality assurance of data supply*

[**ISO 19156**](https://www.iso.org/standard/32574.html) *Geographic information - Observations and measurements*

[**SensorML**](https://www.opengeospatial.org/standards/sensorml) : *describing sensors and measurement processes*

# Recommendations for next generation data models

## Research questions

Before making final recommendations for next generation climate data models, we propose a number of research questions that must be investigated:

* Flexibility vs efficiency – measure the implication of the transposing data stored in “long format”
* Time series data retrieval, with and without [hypertables](https://docs.timescale.com/v0.9/introduction/architecture)

## Data Model Considerations

* Primary keys and indexing
* See [#9](https://github.com/opencdms/opencdms-data-model/issues/9)
* Composite natural keys vs synthetic keys.
* The web: Potential complications introduced by [RESTful APIs](https://en.wikipedia.org/wiki/Representational_state_transfer) and [Object Relational Mapping](https://en.wikipedia.org/wiki/Object%E2%80%93relational_mapping) (ORMs) requiring unique (single) keys.

### Date period

* See [#11](https://github.com/opencdms/opencdms-data-model/issues/11)
* Long vs wide
* See [#10](https://github.com/opencdms/opencdms-data-model/issues/10) - Discussion of normalization, optimization for common scenarios and analysis that requires “[tidy data](https://en.wikipedia.org/wiki/Tidy_data)” (3rd normal form)
* Dynamical schema modifications
* [See #7](https://github.com/opencdms/opencdms-data-model/issues/7)
* MCH allows variations in the database definition. Example include:
* a) Support for table and field names in multiple languages (e.g., Spanish and English)  
  b) Creation of a set of new database tables for each new parameter that is added

# References

|  |  |  |
| --- | --- | --- |
| **#** | **Title** | **Edition** |
| WMO-TD No. 60 | Guidelines on Data Management | 2007 |
| WMO- No. 1192 | WIGOS Metadata Standard | 2019 |
| WMO- No. 1238 | High-quality Global Data Management Framework for Climate | 2019 |
| WMO- No. 1131 | Climate Data Management System Specifications | 2014 |
| WMO- No. 8 | Guide to Instruments and Methods of Observation | 2018 |
|  |  |  |
|  |  |  |
|  |  |  |