



WMO STEWARDSHIP MATURITY MATRIX FOR CLIMATE DATA (SMM-CD)

Developed under the High Quality Global Data Management Framework for Climate

Christina Lief (CCI), Ge Peng (NCICS, NOAA/NCEI), Omar Baddour (WMO), William Wright (CCI),
Valentin Aich (GCOS), and Peter Siegmund (KNMI)

The **World Meteorological Organization (WMO)** is a specialized agency of the United Nations with 192 member states and territories. WMO provides the framework for international cooperation for the development of meteorology, climatology, and operational hydrology. WMO is committed to, and continues to, facilitate free and unrestricted exchange of meteorological and related data and information, products, and services.

The **WMO Commission for Climatology (CCI)** inter-programme initiative called **High Quality Global Data Management Framework for Climate (HQ-GDMFC)** aims at making use of high quality climate data needed for developing climate services for policy and decision making in a variety of applications. A key priority of HQ-GDMFC is to harmonise the definitions and processes and to develop a manual to guide collaborative entities on standards and best practices in the field of data management and stewardship.

The **International Workshop on Information Management**, which was convened by WMO CCI and CBS (Commission for Basic Systems), Geneva, Switzerland, in 2017, included a recommendation for a project plan for climate datasets and access. A key conclusion was that a concept of trusted datasets needs to be defined by a process endorsed by WMO. Datasets must meet standards defined by a maturity index approach. Based on these findings, an **International Expert Group on Climate Data Modernisation (IEG-CDM)** meeting was held at the Royal Netherlands Meteorological Institute (KNMI) in De Bilt, Netherlands in 2018 to develop a climate data-specific version of the maturity model, to be used to assess the “trustworthiness” of climate datasets. Consequently the Stewardship Maturity Matrix for Climate Data Working Group within the IEG-CDM developed the **WMO Stewardship Maturity Matrix for Climate Data (SMM-CD)** self assessment tool. **Subject Matter Experts can use this tool to evaluate 12 aspects, grouped in 4 categories, each scored from level 1 to 5 on the data stewardship of their dataset along with justifications. On this poster for each aspect the five level descriptions are provided.** This provides users with information on the quality of the dataset and a measure of trustworthiness. As the first phase, 18 well-utilized global climate datasets identified by IEG-CDM have been assessed and the SMM-CD assessments are currently under review by the WMO CCL Expert-Team for Data Development and Stewardship.

The members of the SMM-CD Working Group (in alphabetical order)

AICH, Valentin; vaich@wmo.int
BADDOUR, Omar; obaddour@wmo.int
LIEF, Christina; christina.lief@gmail.com
PENG, Ge (Lead); ge.peng@noaa.gov
WRIGHT, William (Co-Lead); wjwright1385@gmail.com

The members of IEG-CDM (in alphabetical order)

1. AICH, Valentin (GCOS)
2. BADDOUR, Omar (WMO)
3. BERGERON, Cedric (ECMWF)
4. BEROD, Dominique (WMO)
5. BUSSELBERG, Thorsten (DWD)
6. CAZENAVE, Anny (LEGOS)
7. DUNN, Robert (Met Office/Hadley Center)
8. GALLAHER, David (NSIDC)
9. GATES, Lydia (DWD)
10. LIEF, Christina (CCI)
11. MILAN, Anna (NOAA/NCEI)
12. PENG, Ge (NCSU/CICS-NC, NOAA/NCEI)
13. ROBERTS, Kate (BOM)
14. SIEGMUND, Peter (KNMI)
15. VERVER, Ge (KNMI)
16. WRIGHT, William (CCI)
17. ZIESE, Markus (DWD)

Categories, Aspects, and Maturity Scales

SMM-CD Category				
Level 1	Level 2	Level 3	Level 4	Level 5
AD HOC	MINIMAL	INTERMEDIATE	ADVANCED	OPTIMAL
Not Managed	Limit-Managed	Managed	Well-Managed	Level 4 +
Not Defined		Defined	Well-Defined	
		Partially Implemented	Fully Implemented	
				Measured, Controlled, Audited

Aspect	Data Access	Usability & Usage	Quality Management	Data Management
Discoverability	Data Portability	Quality Assurance & Control	Preservation	
Accessibility	Documentation	Quality Assessment	Metadata	
	Usage	Uncertainty Analysis	Governance	
		Data Integrity		

Acknowledgement

This work was originated at the International Expert Group on Climate Data Modernisation (IEG-CDM) meeting held by WMO at the Royal Netherlands Meteorological Institute (KNMI) in De Bilt, Netherlands on April 16–18, 2018. Travel support was provided by WMO and/or participant’s home countries. Dominique Berod, Robert Dunn, David Gallaher, Lydia Gates, and Markus Ziese contributed to v00r01 20180418. Comments from the IEG-CDM members and Iolanda Maggio, Peter Thorne, Simon Eggleston, Darren Ghent, Jörg Schulz, Nancy Ritchey, Ken Kehoe, Imke Durre, Carolin Richter, Ruth Duerr, and Axel Andersson were beneficial.

SMM-CD CATEGORIES AND EXPECTED BEHAVIOURS FOR EACH ASPECT

Data Access Category

This category refers to the ability to locate (Discoverability) and get to the dataset in question (Accessibility), with higher levels of maturity corresponding to the ease for a potential user to find and gain access to the dataset.

Aspect	Level 1	Level 2	Level 3	Level 4	Level 5
Discoverability	By personal contact only; Dataset information not discoverable	Limited dataset information, such as scientific description of the methodology, in the literature	Minimal catalog-level metadata; Dataset searchable online	Complete set of collection-level discovery metadata + minimal granular metadata	Level 4 + available on an international catalogue, prominently displayed online and routinely updated
Accessibility	Data not available publicly; Person-to-person contact needed	Basic online services available for data access (e.g. FTP/HTTP direct download).	Non-standard data service	Standard-based interoperability data service	Level 4 + full capability of sub-setting, aggregation and visualization

Quality Management Category

Quality management encompasses quality assurance procedures including quality monitoring, quality control, and quality assessment and communication of reliability.

Aspect	Level 1	Level 2	Level 3	Level 4	Level 5
Quality Assurance & Control	Data quality assurance (QA) & control (QC) procedure unknown or none	QA/QC procedure are defined, documented, and partially implemented	QA/QC procedure are well-defined according to community best practices, documented and fully applied	Level 3 + provision of error statistics published or tracked with results made available online and communicated to data providers; Procedure for user feedback, improvement prioritization in place	Level 4 + detailed analysis of errors and gaps at space-time unit level: (Station, grid-points, daily, monthly and or annual time-scale, etc.) QA/QC procedure monitored; Retrospective QC
Quality Assessment	Product quality assessment not done or done internally and information not available	Assessed by Principal Investigator (PI) or data producer; Assessment results available online	Level 2 + Product validation and evaluation done by PI published in peer-reviewed journal.	Level 3 + Independent product validation and evaluation published in peer-reviewed journal.	Level 4 + the complete product provenance is captured and publicly available
Uncertainty Analysis	Uncertainty estimates not available.	Uncertainty estimates presented without explanation.	Uncertainty estimates presented with partial explanation.	Full uncertainty budget available with all assumptions; Estimates of accuracy of trend available.	Full uncertainty assessment published in peer reviewed journal.
Data Integrity	Unknown or no data integrity check	Random data integrity check	Data integrity verified systematically but methodology not commonly known	Data integrity systematically verified and following well known practices but not necessarily consistent across platforms	All steps in data integrity check systematically verified and adhering to well-known practices and reported

Usability and Usage Category

This category describes how easily the data product may be understood and used by users and incorporated into the user’s own working environment.

Aspect	Level 1	Level 2	Level 3	Level 4	Level 5
Data Portability	Non-machine readable	Basic machine readable	Standards-based machine readable	Machine independent, self-describing, interoperable format	Level 4 + capability of providing user required format
Documentation	Product information not publicly available online	Limited online documentation (e.g., User Guide)	Document on how the data product was created and how to use it, is available online	Full documentation based on a standard template and available online	Level 4 + online tutorial on using and analysing the dataset; Complete production system information available online
Usage	No or weak citations in scientific publication in peer-review journal or as institutional reports	Intermediate citations + referenced in institutional climate assessment reports (e.g., by NOAA)	Strong citations + referenced in national climate assessment reports (e.g., by USGCRP)	Level 3 + referenced in international climate assessment reports (e.g., by IPCC)	Level 4 + referenced in international decision/policy making published reports (e.g., by UNFCCC, UN-ISDR, World Bank, etc.)

Data Management Category

This category refers to the processes undertaken to ensure the data and the “contextual” metadata are securely archived. It covers not just the preservation of the data and metadata with appropriate safeguards, but well defined and enforced governance processes to ensure that the right procedures are followed at the right times by the right people.

Aspect	Level 1	Level 2	Level 3	Level 4	Level 5
Preservation	Any storage location; Data only; Data not backed up	Non-designated repository; A backup copy of electronic data is made	Designated archive; Basic retention policy defined. Routine backups made, including offsite copy	Level 3 + conforming to community archiving standards. Comprehensive retention policy defined and implemented	Level 4 + archiving process performance controlled, measured and audited; Future archiving standard changes planned
Metadata	Metadata not publicly available and/or not usable	Limited Metadata publicly available; Conforming to community-standard; Basic characteristics of dataset	Level 2 + conforming to international standards in most aspects; limited quality and provenance metadata	Fully compliant with international standards; Rich metadata content; Basic granular-level metadata; Support dataset provenance	Level 4 + complete granular-level metadata; Metadata QC-ed and Regularly updated
Governance	Responsibility is not defined; No person is assigned.	Responsible entity is identified; Accountability and competency are not well-defined.	Responsibility/accountability, and compliance mechanisms are defined; Good competency; Processes established conforming to community standards	Level 3 + competency defined; Confirming to international standards; auditable	Level 4 + accountability and responsibility well-defined and fully compliant with international standards; transparent; Monitored and audited

