|  |  |  |
| --- | --- | --- |
| WEATHER CLIMATE WATER | **World Meteorological Organization**  **COMMISSION FOR OBSERVATION, INFRASTRUCTURE AND INFORMATION SYSTEMS**  **Second Session** 24 to 28 October 2022, Geneva | **INFCOM-2/INF. 6.3(1.3)** |
| Submitted by: Secretary-General  27.IX.2022 |

CLIMATE DATA MANAGEMENT AND THE WMO INFORMATION SYSTEM 2.0

### 1. OpenCDMS – Development and Progress

Introduction

Effective management of meteorological, hydrological, oceanographic and other environmental data (referred to below as Earth system data) is critical for the monitoring and analysis of the current state of the Earth system and for the seamless prediction of its future state. This monitoring, analysis and prediction, in turn, is essential for the effective and timely provision of related services, including climate services.

Underpinning the effective management of Earth system data is the use of modern Climate Data Management Systems (CDMS). In this context the term climate data is intended to include any Earth system data, metadata and information of sufficient spatial coverage, temporal length, consistency and continuity to determine climate variability and change. This includes: observational data (in situ and remotely sensed); derived products and analyses; and the output from numerical predictions or simulations.

Timeline

Description automatically generated with medium confidence

**Figure 1: Roadmap for the implementation of a reference Climate Data Management System**

In recognition of the importance of effective CDMSs the WMO published the WMO [*Climate Data Management System Specifications*](https://library.wmo.int/index.php?lvl=notice_display&id=16300) (WMO-No. 1131), establishing a framework defining the functionality required within a climate data management system at a high level. The framework aims to promote the standardization and interoperability between climate data management systems, improve the availability and timeliness of climate data and the adoption of best practices by Members, including the use of a climate data management system where one is not already in use.

Noting the importance of effective climate data management, the eighteenth World Meteorological Congress requested the development of a reference open-source climate data management system to support members and to provide a reference implementation of a climate data management system that meets the specifications. The OpenCDMS project was subsequently established according to the timeline for development as shown in Figure 1. This document provides a summary of progress towards those goals and the implementation of a reference system.

Graphical user interface

Description automatically generated

**Figure 2: Functional layers under development as part of the OpenCDMS project**

Architecture

The reference climate data management system under development as part of the OpenCDMS project follows the layered approach shown in Figure 2. The layers shown are:

* Data layer – storage layer providing a reference implementation of the climate data model standard.
* Logic Layer – layer that implements the core business logic specified within the WMO CDMS Specification (WMO-No. 1131) and that provides a standardized web API for accessing and managing the data.
* Presentation layer – Web application and interface providing access to the observations and derived products, including associated metadata. This includes components for data entry, quality control and calculation of standard WMO products.

This system has been designed so that it can be installed by Members who either do not currently use a CDMS or who need to update to a more modern system.

Implementation and progress

Implementation has focused on developing a component from each layer shown in Figure 2. Due to delays with the climate data model (see Section 2), and due to funding sources and constraints, initial development of the data layer has focused on building support for an existing CDMS, Climsoft.

Development of the presentation layer (*opencdms-app*) has focused on metadata management, data entry and an interactive form builder. Example screenshots for a demonstration version of the application are shown in Figures 3 and 4 respectively. Figure 4 shows the data management page for stations using example historic data provided as part of the Climsoft v4 database repository (<https://github.com/climsoft/climsoft-database>). Figure 5 shows a data entry form for synoptic weather reports.

The demonstration web application can be found at: <https://demo.opencdms.org/>

Access control is implemented as part of the demonstration web application and access can be gained by request from [info@opencdms.org](mailto:info@opencdms.org). Feedback can also be provided to this address and is welcomed.

Graphical user interface, application

Description automatically generated

**Figure 3: Example screenshot from the OpenCDMS demonstration web application configured with support for a Climsoft database. The image shows the list of available stations in the metadata management component.**

A screenshot of a computer

Description automatically generated

**Figure 4: Screenshot showing example data entry form synoptic weather reports from the Caribbean region**

Outlook and future work

Parallel to the work within the OpenCDMS project, and to meet their national needs, the National Meteorological Service of Belize has developed a climate data management system (Surface), using the WMO CDMS specification (WMO-NO. 1131) as a template. This system is now used operationally in Belize and has recently been shared with the international community, with the source code made available through the OpenCDMS code repository (<https://github.com/opencdms/surface>). Given the mature state of the Surface CDMS, ongoing work within the OpenCDMS project is focused on a convergence between the Surface CDMS and the OpenCDMS project. Where applicable, components from the data and logic layers of the OpenCDMS project will be integrated into Surface to develop a new production ready system, OpenCDMS Surface. Integration of the OpenCDMS Surface into the WMO Information System (WIS) 2.0 will also be developed as part of the WIS 2.0 pilot project (See Section 3). The provisional OpenCDMS Surface, together with the proposed OpenCDMS reference implementation architecture, will be presented to Congress-19. Testing of the system will begin in 2024 and with an installable, production ready system aimed to be available during 2025.

### 2. Task Team – Climate Data Model (TT-CDM)

As noted within the WMO [*Climate Data Management System (CDMS) Specifications*](https://library.wmo.int/index.php?lvl=notice_display&id=16300) (WMO‑No. 1131), the use of a standard data model for the representation and exchange of climate data will facilitate the interoperability between climate applications and components by providing a climate data management system independent view of the data. Once established the data model will:

* Help establish standard vocabularies for climate data and related information
* Harmonize climate data management practices among WMO members
* Facilitate effective exchange of climate data via an independent logical data model and associated serialization
* Underpin future climate services

This development and publication of a data model is also fundamental to the implementation of a reference open-source climate data management system and the OpenCDMS project as described in Section 1.

In recognition of this importance, the Task Team – Climate Data Model (TT-CDM) was established in early 2022 as part of the Expert Team on Metadata Standards (ET-Metadata) with the following terms of reference:

* Develop and maintain a WMO standard data model for the representation, exchange and archival of climate data in support of Climate Data Management Systems (CDMSs) in line with the WIGOS Metadata Representation and according to the requirements provided by ET-DRC;
* Coordinate with relevant WMO expert teams such as ET-Data, ET-Metadata, ET-IM, to align the data model with existing WMO standards;
* Support the OpenCDMS project in the implementation of the data model ([Resolution 22 (Cg-18)](https://library.wmo.int/doc_num.php?explnum_id=9827/#page=95)); [Resolution 21 (EC-73)](https://library.wmo.int/index.php?lvl=notice_display&id=22032/#page=363).

Once developed, the data model is expected to be adopted as a WMO standard that will become the model used for the management and exchange of climate data by Members and WMO centres, as required under the WMO Unified Data Policy. This is in addition to use within climate data management systems.

Activity in the Task Team has focused on identifying requirements and foundational standards for the data model, building on the previous work within the Expert Team on Data Requirements for Climate Services (ET-DRC) and [*Climate Data Management System Specifications*](https://library.wmo.int/index.php?lvl=notice_display&id=16300) (WMO-No. 1131). As part of this work, the draft Open Geospatial Consortium (OGC) *Observations, Measurements and Samples* (OMS) standard has been identified as the foundational standard. The OMS standard is expected to be published in the second half of 2022 or early 2023 as ISO 19156:2022, *Geographic information — Observations, measurements and samples.* The WMO Core Metadata Profile (<https://community.wmo.int/WIS2_Technical_Specification_Guidance>) and the [*WMO Integrated Global Observing System (WIGOS) Metadata Standard*](https://library.wmo.int/index.php?lvl=notice_display&id=19925) (WMO-No. 1192) will form the other foundational building blocks for the climate data model. Ongoing work includes identifying gaps in the data models and defining the extensions required for climate data and for full traceability of the processing applied to the data.

### 3. Integration of the OpenCDMS into the WIS 2.0 framework

Climate data and information can be exchanged using the current WIS infrastructure. Observational daily (DAYCLI) and monthly (CLIMAT) climate summary reports are discoverable, accessible and retrievable from the WIS. Similarly, it is possible to discover, access and retrieve information such as seasonal outlooks and summaries. However, whilst this is possible, there are a number of barriers to both the exchange and publication of the data and their discovery, access and retrieval. These barriers, or issues, are related to:

1. The creation and publication of the metadata describing a resource;
2. The granularity of the discovery metadata records when searching;
3. Inhomogeneity of access and retrieval methods;
4. Transformation of the data to the required formats, typically Binary Universal form for the Representation (BUFR) for observations.

These issues have been widely recognized and are applicable to other domains and application areas. Consequently, the WMO Information System is undergoing an evolution as the WIS 2.0, seeking to address many of the issues. This will be achieved through the simplification of the discovery metadata, improving the granularity of the metadata records and the use of open Web standards and services for the discovery, access and retrieval of data and information. As part of the WIS 2.0 infrastructure global services will be provided for:

* Discovery of data through a global catalogue (global catalogue)
* Publication, subscription and notification of new data (global message broker)
* Access and retrieval of data (global cache)

These global services will be provided on resilient, scalable infrastructure to ensure continuity of service and meet the performance requirements (e.g. low latency, high bandwidth). As part of the WIS 2.0, individual nodes (WIS2.0 nodes) will connect to these services to synchronize discovery metadata, publish and subscribe to notifications of new data and to access and retrieve data relevant to local requirements. On receipt of notification of new data from a WIS 2.0 node the global broker forwards the message to subscribers, including the global cache. The data is then retrieved by the global cache and made available to global users. The global services, architecture and key functions are described further in [INFCOM-2/INF 6.3.1(4)](https://meetings.wmo.int/INFCOM-2/InformationDocuments/Forms/AllItems.aspx).

To support the development of the WIS 2.0, a reference implementation of a WIS 2.0 node has been developed and is described in [INFCOM-2/INF 6.3.1(1)](https://meetings.wmo.int/INFCOM-2/InformationDocuments/Forms/AllItems.aspx). For climate data, the key functions and underlying standards of the reference implementation are:

* Publication of discovery metadata (OGC API – Records)
* Publication and subscription to notification and data services (MQTT)
* Provision of data through open web APIs (OGC API – Features)

The reference implementation also contains functionality to transform data from incoming text files to the WMO BUFR format to address issue (4).

In order to test and demonstrate the interoperability of the OpenCDMS project and integration of climate data in the WIS2.0 framework a project focused on climate data is planned for the WIS2.0 pilot phase. This project will develop and implement an automated workflow for the extraction, transformation and loading of daily (DAYCLI) and monthly (CLIMAT) climate data from a CDMS (Surface CDMS, Belize) into a WIS 2.0 node and subsequent publication on the WIS 2.0 infrastructure. This process will include the publication of discovery metadata, conversion to BUFR, notification of new data via the WIS 2.0 publication / subscription protocols, and the provision of data through a Web API. Parallel to this, the CDMS system will subscribe to access relevant real time data available through the WIS 2.0 for ingestion into the CDMS. The key improvements will be:

* Improved discoverability and accessibility of the climate data on the WIS 2.0 infrastructure
* Improved timeliness and consistency of the climate data exchanged via the WIS 2.0
* Improved availability of the real time data within the CDMS for further processing and inclusion in the climate records

Whilst the pilot project will focus on the DAYCLI and CLIMAT data it should be noted that the processes and metadata are equally applicable to the full climate record and across Essential Climate Variables (ECVs) and domains (e.g. rescued and digitized data archives, long term historical holdings, seasonal forecasts etc.).

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_