Project Proposal
Angola Data Rescue Pilot Project:
Ingestion in the INAMET CDMS of
digital historical records of meteorological observations

Table of Contents

1. Background of this undertaking

WMO, with support from the EU-funded Climate Services and Related Applications Programme (ClimSA), is undertaking an initiative during the period December 2023 through mid-2024 to support the Southern African Development Community Climate Services Centre (SADC CSC), along with the respective ClimSA regional focus country, Angola, in a variety of data-related areas including data rescue, climate data management system (CDMS) implementation, and data exchange.

The Project Proposal for the Angola Data Rescue Pilot Project, below, has been developed by a group of colleagues/experts representing:

- Angola National Institute of Meteorology and Geophysics (INAMET)
- Southern African Development Community Climate Services Centre (SADC CSC)
- the WMO SERCOM Expert Team on Data Development and Stewardship (ET-DDS)
- the WMO INFCOM Expert Team on WIGOS Tools and Regional WIGOS Centres (ET-WTR)
- Royal Netherlands Meteorological Institute (KNMI)
- International Environmental Data Rescue Organization (IEDRO)
- Innovations in Development Education and the Mathematical Science International (IDEMS)
- Meteo-France International (MFI)

- German National Meteorological Service (DWD)
- the WMO Secretariat.

This group is informally known as "The Angola Data Rescue Group". This initiative was technically coordinated on behalf of WMO by Lisa-Anne Jepsen, WMO Consultant (ljepsen.wmo@gmail.com).

2. Status of data rescue in Angola: Unified inventory of rescued data

Thanks to data rescue efforts undertaken by INAMET and partners, including the present WMO-led data related initiative, INAMET now has available a summary of inventory observations and periods of record from international databases in the United States, Europe, and the Southern Africa region. This unified inventory indicates worldwide availability and location of digital historical records for Angola's meteorological observations (ref: ANGOLA.Unified Inventory, data rescue, rev 2may2024) . Included in this workbook are the sources of both digitized data and non-digitized data (imaged paper records that have yet to be digitized).

With support from various climate data rescue initiatives including that implemented by FAO (2018-2021), and by SASSCAL (sponsored by the German Federal Ministry of Education and Research and supported by DWD, 2013-2017), INAMET Angola has imaged its climate data that were formerly available only on paper records. INAMET has also benefited from the recently conducted Modernization Project that supported INAMET in implementing the complete meteorological value chain from basic infrastructure to service delivery targeted at end-users. The modernization project relied on WMO-compliant components and was technically supported by MFI.

Approximately 200,000 pages of very dense information may now require digitization (manually keying into the computer) in order that INAMET finalize its data rescue project. The cost of digitizing all these pages is very high and could easily surpass hundreds of millions of dollars.

However, much of this information may have already been digitized through projects supported by the European Union and others. Thus, INAMET must now determine the set of meteorological observations that still require final digitization and compare that set to the imaged paper records. This analysis will significantly reduce the amount of costly digitization to complete the Angola data rescue project.

To automate this comparison of climate datasets and produce standard statistical products, INAMET can use functions available in a climate database management system (CDMS) that meets WMO specifications (ref: Climate Data Management System Specifications, WMO-No. 1131). Using its CDMS, INAMET can compare digitized information to imaged records and identify gaps in the observation periods of record. But to support such a comparison of climate datasets, the CDMS must include all the digital information available from the various international databases holding that digital information.

3. Objectives: benefits for INAMET and the international climate community of the Angola Data Rescue Project

Once all of Angola's climate data have been ingested in a single CDMS, INAMET will be positioned to realize the following benefits:

1.) Identify remaining imaged records that require digitization to complete the historic record.
   a. Compare digital vs imaged records, using the "query" functions in the CDMS; this is an efficient and cost-effective method of comparison

   b. Increase INAMET's technical capacity in database management and use thanks to having undertaken the exercise of ingestion and comparison

2.) Avoid re-digitizing imaged records that are already digitally available thanks to various international data rescue projects, thereby potentially saving hundreds of thousands of dollars in manual digitization costs by only digitizing imaged records that complete the historical record.

The international climate community, including INAMET, will realize the following benefits:

3.) Access and use the totality of Angola's digitized meteorological observations, supporting and improving:
   a. Forecasting and numerical modeling capabilities
   b. Disaster mitigation, warning, and planning
   c. Land use and management decisions

4.) Learn from/build on the example of the Angola Data Rescue Project

4. Proposed technical approach and training, Angola Data Rescue Pilot Project

The Angola Group was invited by WMO to formulate a project proposal for an Angola Data Rescue Project.

Proposed technical approach
To test the technical approach, the Angola Group proposes implementing a *Pilot Project* (Phase I of the *Extensive Angola Data Rescue Project*). With the Pilot Project, the Angola Group aims to **test a proposed technical approach by which INAMET shall ingest in its CDMS historical data that have already been digitized and are available in tabular form**.

The technical approach for the Pilot Project is summarized below in the *Diagram: Technical approach for the Pilot Project* and detailed in *Appendix 1 - Proposed technical approach, Angola Data Rescue Pilot Project*.

Aspects the Angola Group have borne in mind and that are further discussed in Appendix 1 include: the need for traceability and security of the data from the various sources, the need to make use of the technical capacities of the INAMET CDMS, CliSys, and the need to strengthen the capacity of INAMET staff members to make best use of Clisys and other data-related tools and applications including SQL.

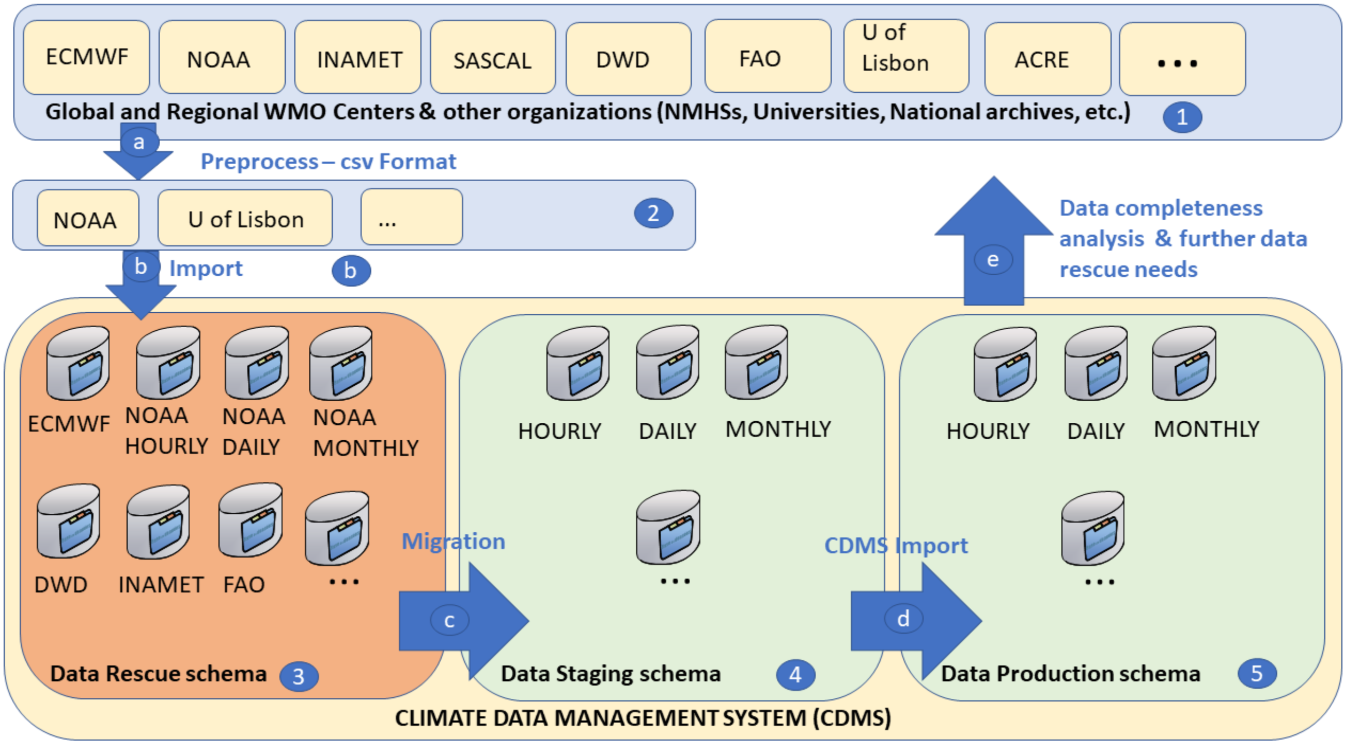
MFI has pointed to the following caveat: INAMET has contracted MFI to maintain Clisys. This contractual framework is not compatible with a customization/use of the database from a third party. Therefore, any database schema adaptation/extension to Clisys will require a technical study by MFI to evaluate the feasibility, the risks and the associated workload. The Pilot Project will thus implement, to the extent possible, the Steps and Processes that do not require enhancements to the current Clisys configuration. Technical studies, if required, may be carried out by MFI under the Extensive Angola Data Rescue Project.

The technical approach for the Pilot Project includes the following steps:

Step 1: Historical climate data inventory and digitization
Step 2: Import the preprocessed files into the CDMS Data Rescue Schema
Steps 3, 4 and 5: Data rescue, Data staging and Data production within the CDMS

- **Steps 1, 2, 3, 4, and 5** - each step results in a stable situation, that is, an intermediary result; and
- ***Processes a, b, c, d and e*** - each process contributes to the accomplishment of a step.

Diagram: Technical approach for the Pilot Project



Experience in migrating historical data to Clisys indicates that Steps 1-5 are an iterative process: input data and files are typically quality controlled and modified several times before being used in production mode.

Training

The proposed methodology for each of the Steps listed above sees a combination of support provided by expert bodies/individuals and training for INAMET.

Training will include practical exercises that take up the digitization, data rescue, CDMS integration/operational tasks under the guidance of an expert body/individual. In the case of each step, the expert body will be required to provide detailed manuals/documentation on process employed and on training topics; INAMET is to participate in the development of these manuals to ensure that the manuals address all required tasks.

Training and support topics shall include data-rescue related functions such as: digitization of rescued data; use of the INAMET CDMS for data rescue; use of a RDBMS; use of statistical tools, applications and programming languages pertinent to data rescue/management; these may include: R-Instat, SQL, and Python.

The Angola Data Rescue Project foresees delivering to INAMET a manual on data rescue standard operating procedures (SOP) for Angola that will serve as a *vade mecum* for future such endeavours.

5. Pilot project timeline, deliverables, inputs, implementation modalities, and budget

Timeline: 14-month period, November 2024- December 2025. A detailed Gantt chart that includes due dates for the deliverables below will be developed as part of the inception report for the Pilot Project.

Deliverables:

- Digitized rescued data ingested in the INAMET CDMS and operational; these data shall include two sources contained in the Unified inventory of rescued data, Angola (NCEI, NOAA[1] and University of Lisbon). For the pilot project only the daily data from both sources will be imported into the CDMS.
  - Preprocessed data from the two sources: NCEI, that is already in the CSV format (ref: tab *NCEI Sample AO-6* in the Unified inventory of rescued data, Angola); and from University of Lisbon, that is not in the CSV format (ref: tab *U Lisbon, Sample AO-2* in the Unified inventory of rescued data, Angola).
  - Data ingested in the CDMS: NCEI and University of Lisbon data ingested in the CDMS
  - Data operational in the CDMS: NCEI and University of Lisbon data operational in the CDMS
  - Analysis of data completeness and possible needs for further data rescue

- Enhanced INAMET capacity to use the CDMS and related tools for preprocessing/ingestion/operationalization of rescued data: INAMET staff members, including staff from both the IT Department (in particular, the INAMET Database Manager) and the Climate Department, and short-term technicians trained on topics including:

  - Preprocessing of climate data from Excel files to CSV format (using Shell scripts, program languages, specific applications; this process is independent of Clisys)
  - Assuring data quality and traceability within the CDMS; (this process is handled natively in Clisys for CSV ingestion)
  - Manipulating digital data from various sources to launch/continue the process of data ingestion from all available sources (in addition to the two sources contemplated within this Pilot Project; this process is independent of Clisys)
  - Adding more historic data to the CDMS as these data become available: both manipulation of historic data and ingestion of historic data in the CDMS[2]
  - Accessing data to enhance INAMET's ability to use its expanded database that includes not just current, but historic data

- New data made available to the scientific community
  - Send data or make data from the University of Lisbon available to global centers (NOAA, DWD, ECMWF, etc.), and update data on the WMO data rescue portal. That will participate in the integration of these data to upcoming reanalysis or numerical projections.

[1] National Centers for Environmental Information (NCEI), National Oceanographic and Atmospheric Agency (NOAA) is referred to in this document as "NCEI"

[2] Technical observation by MFI: the most time-consuming aspect is the creation of Clisys parameters, allowing values to be stored in the CDMS. Given that the Clisys web interface is not designed for a massive creation of parameters, an intervention by MFI will likely be necessary.

- o General methodology to be proposed to NMHSs for inserting historical data into their system in guaranteeing data traceability and data security

See *Appendix 2 - Supplemental information on deliverables and managing system security* for more detailed information regarding the Pilot Project deliverables.

Inputs:

Human resources:

In-kind human resources contributed by INAMET:

A. INAMET Coordinator of the Implementing Arrangement with WMO (Isildo Ntemo Gomes)
B. INAMET Database Manager (Edson Segunda)
C. INAMET Climate Department interface (António Manuel Lameira Gaspar)

External human resources whose cost shall be funded by The Pilot Project:

A. External Manager of the Pilot Project: day-to-day management of the technical and administrative aspects of the Pilot Project including integration of the outputs of the Implementing Arrangement between WMO and INAMET
B. External Technicians (two locally recruited technicians): for a period of approximately one year, to support INAMET in carrying-out Processes a, b, c and d described above; the External Technicians shall closely collaborate with the INAMET IT Department, (*or whichever department administers the CDMS)*
C. Technical Experts to provide guidance and support capacity development to ensure delivery of the deliverables listed above; technical experts to address:
   a. Data development and stewardship
   b. Data conversion
   c. Clisys - definition of the types of CSV files that may be ingested in Clisys; ingestion in Clisys of preprocessed data; import and creation of tables in Clisys including QC; scripts for exporting/sharing data

Hardware:

A provisional list of hardware follows; all hardware procured under the Pilot Project is to be compatible with existing INAMET technical infrastructure and shall not duplicate hardware to be provided by SADC under its ClimSA grant.

- Storage capacity - Hard disk - HDD; 2 discs, 10 TB each (TBC; assumes the volume of data already digitized totals 100-150 GB)
- Laptop computers: 2 PCs (laptop computers) - one laptop for each of the two External Technicians (ref: *External human resources inputs*, above) - a secure PC that is able to connect to the INAMET IT network in order to safely run Clisys in a performant and compatible way

Project implementation modalities:

The Angola Group proposes that the Pilot Project be implemented using two modalities:

1. **Implementing Arrangement between INAMET and WMO** that will provide funding to INAMET to cover the following inputs (ref: *Human resources* section above)
   - A. Locally-recruited External Technicians
   - B. Reporting on the agreed Implementing Agreement deliverables

2**. Direct funding by WMO of external human resources and other inputs** including:
   - A. External Manager of the Pilot Project
   - B. Technical Experts, guidance and training
   - C. Hardware (final list TBC)

Pilot Project Ad-Hoc Advisory Body: The Angola Data Rescue Group shall function as a Pilot Project Ad-Hoc Advisory Body and shall be regularly updated and consulted by the Pilot Project management; the Angola Data Rescue Group includes representatives from among the following organizations:

- Angola National Institute of Meteorology and Geophysics (INAMET)
- Southern African Development Community Climate Services Centre (SADC CSC)
- the WMO SERCOM Expert Team on Data Development and Stewardship (ET-DDS)
- the WMO INFCOM Expert Team on WIGOS Tools and Regional WIGOS Centres (ET-WTR)
- Royal Netherlands Meteorological Institute (KNMI)
- International Environmental Data Rescue Organization (IEDRO)
- Innovations in Development Education and the Mathematical Science International (IDEMS)
- Meteo-France International (MFI)
- German National Meteorological Service (DWD)
- the WMO Secretariat.

Budget:

Budget, Angola Data Rescue Pilot Project

Nov 2024-Dec 2025

	unit	**number of units**	**cost per unit in EUR**	**cost per item in EUR**
Implementing Arrangement - INAMET and WMO				
External Technicians - 2 technicians; 10 months ea.	month	24	500	10000
Reporting costs	project	1		1000
Total, Implementing Arrangement				**11,000**
Direct funding by WMO				
Human resources				
External Manager of the Pilot Project (10-20% of full time position)	project	1		0
Technical Experts				
Expert, Data Development and Stewardship	project	1		0
Data conversion- Guidance and training	project	1		0
Clisys - training and support in: defining the types of CSV files that may be ingested in Clisys; ingestion in Clisys of preprocessed data; import and creation of tables in Clisys including QC; scripts for exporting/sharing data	project	1		0
Total human resources				**0**
Hardware		1		
Storage capacity - Hard disk - HDD; 2 discs, 10 TB each (TBC; assumes the volume of data already digitized totals 100-150 GB)	HDD	2	1000	2000
Laptop computer	computer	2	1300	2600

Total hardware				**4600**
Total, Directly funded by WMO				***0***
Grand total budget - EUR 40,000				***8,400***

6. International Data Rescue Portal

WMO recommends that INAMET add the Angola Data Rescue Project to the International Data Rescue Portal that is operated by Copernicus in collaboration with WMO. The International Data Rescue Portal is a successor to the IDARE Portal. A presence in the International Data Rescue Portal will help make the Angola data rescue efforts discoverable by the data rescue community and potential donors that may wish to collaborate with Angola in the future.

7. List of guidance documents

References that will guide the implementation of this initiative include those saved here: WMO ClimSA data-related resources.

8. Further support by WMO Secretariat

The WMO Secretariat stands ready to provide technical support and guidance in the implementation of the Next Steps identified in this *Summary Road Map for a Samoa Data Rescue Project*. For further information, the Samoa Meteorological Service and other actors are invited to contact:

Mr Omar Baddour
Head Climate Monitoring and Policy Services Division (CMP)
Climate Services Branch
Services Department
World Meteorological Organization ¦ 7bis, avenue de la Paix, 1211 Geneva 2, Switzerland
Office: +41 22 730 82 68
E-mail: obaddour@wmo.int

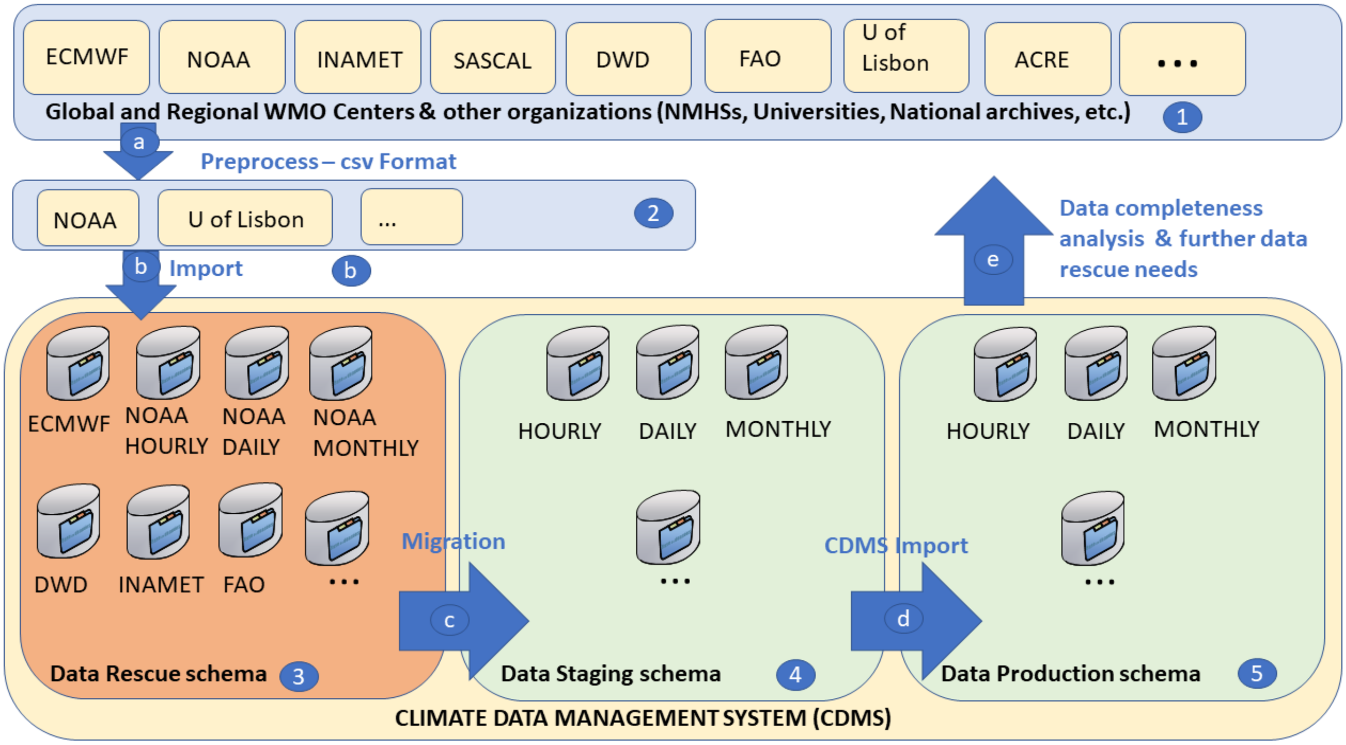
Appendix 1
Proposed technical approach, Angola Data Rescue Pilot Project

The proposed technical approach for the Pilot Project includes the following steps and processes that are shown in the diagram below:

- **Step 1**: **Historical climate data inventory and digitization**
- **Step 2: Import the preprocessed files into the CDMS Data Rescue Schema**
- **Steps 3, 4 and 5: Data rescue, Data staging and Data production within the CDMS**

- **Steps 1, 2, 3, 4, and 5** - each step results in a stable situation, that is, an intermediary result; and
- ***Processes a, b, c, d and e*** - each process contributes to the accomplishment of a step.

Diagram: Technical approach for the Pilot Project



Past experience in migrating historical data to Clisys indicates that Steps 1-5 are an iterative process: input data and files are typically quality controlled and modified several times before being used in production mode.

Step 1: **Historical climate data inventory and digitization -** The proposed technical approach for the Pilot Project builds on work accomplished prior to the start of the Pilot Project under Step 1 and that resulted in the following deliverables (ref: blue highlighted Box 1 in the *Diagram: Technical approach for the Pilot Project*)**:**

- identification of all potential data to be rescued for Angola: a unified inventory of all rescued data; this inventory reflects input from all available sources including NCEI, C3S, DWD, U of Lisbon, etc. and includes:
    - **digitized data** (that is, data in table form that reflect the contents of images that were scanned and saved electronically) and
    - **non-digitized data** (that is, images that have been scanned and saved electronically but whose contents have yet to be converted into digital tables)
- specification of the meta-data regarding the digitized and non-digitized data that available (e.g., source of the data; parameters; quality of the data; contact information for the data)
- digitization of images of priority rescued data and tables (for example, in Excel) that reflect the contents of these images[3]

The Pilot Project builds on Step 1 (described above) and sees the following steps and processes:

- preprocessing of two examples of digitized climate data;
- ingestion of the digitized, preprocessed climate data into the INAMET CDMS [4], and operationalization of the data within the CDMS; and
- comparison of digital inventory and priority data to identify images that have yet to be digitized

Process a: Preprocess the data - convert the Excel files of digitized data to CSV format **-** Data must be in the CSV format to be ingested in the CDMS. The Unified inventory of rescued data, Angola includes examples of both data already in the CSV format (ref: tab *NCEI Sample AO-6* in the Unified inventory of rescued data, Angola); and data not in the CSV format (ref: tab *U Lisbon, Sample AO-2* in the Unified inventory of rescued data, Angola).

The Pilot Project will demonstrate approaches for preprocessing data contained in the *Unified Inventory of rescued data, Angola.*

[3] Regarding the digitization of images of rescued data:

- the Pilot Project for Angola Data Resuce will make use of data that have already been digitized and are available in tabular form; and
- for the Extensive Angola Data Rescue Project, INAMET has requested that, future digitization make use of a process of character recognition that could transform scanned images into numerical values.

[4] Clisys requirements regarding ingestion of custom CSV files include:

- Previously to the import
* parameters are created in Clisys
* storage table(s) with these parameters are created
- CSV values are numerical only (integers or floats)
- CSV structure is vertical (one column per parameter)
This ingestion process goes through a first level of quality control rejecting absurd values. Therefore, the stored data cannot be the exact image of raw input data, which would mean that traceability would be lost .
Once in the database, data can be manipulated using Clisys features including: visualization, quality control, computations

Step 2: Import the preprocessed files into the CDMS Data Rescue Schema - The files that have been preprocessed (ref: ***Process a***) are imported into the CDMS Data Rescue Schema.

Technical observation by MFI: In Clisys, each parameter is a database column and thus MFI strongly recommends creating dedicated parameters (and so dedicated columns) for rescue data to avoid that INAMET accidentally overwrites existing data. For example, a new "daily mean temperature" parameter for data rescue (and data staging) should be created, rather than importing into existing "daily mean temperature" (that is the operational table).

Processes b, c and d: Ingest the CSV data into the CDMS - The description and diagram below detail a proposed strategy for ingesting into the CDMS the rescued data in CSV format. This strategy foresees the creation of 2 database schemas:

Steps 3, 4 and 5: Data rescue, Data staging and Data production within the CDMS

Create 2 schemas in the CDMS (ref: **Steps 3 and 4** as shown in the diagram):

- the **Data Rescue Schema (Step 3)** will store the different data rescue sources (1 table per file) and will ensure the data traceability/lineage,[5] and easy data access[6].). This schema and the "**Preprocess – csv Format**" could then be reused for any other CDMSs when importing such kind of data.
- the **Data Staging Schema (Step 4)** [7] will store the result per data type (hourly, daily, monthly) and reflect values available in all the different data rescue sources. This will be done in a working area (referred to as a 'staging environment area') and will ensure the correct functioning of the operational CDMS processes (ref: the **Data Production Schema, Step 5**).

Process b: import the data from the data rescue sources (that is, data that have been preprocessed and are in the CSV format); This step could be performed by different mixed means: SQL language, shell script, program language, application, database tools; notably, for Clisys a tool is available to manage the CSV import.

[5] Data traceability is ensured for ingested data; rejected data are lost.

[6] Data Rescue Schema -Technical observations by MFI: to preserve data consistency, in Clisys it is not recommended to execute SQL directly in the database; the CSV ingestion process is designed for data import.

[7] Data Staging Schema – Technical observation by MFI: Assumes the Data Staging Schema is compatible with Clisys features; as it is not recommended to manipulate schemas and tables natively in SQL in the Clisys database, the data rescue schema should be inside the Clisys database.

Furthermore, MFI notes: it is not recommended to migrate "manually" the data from Data Rescue to Data staging; re-importing CSV files into data staging is a better idea. Therefore, the ideal workflow could be:

CSV processing tasks coding --> CSV generation ("Data rescue" dataset) --> import into Clisys --> errors and rejection checks --> CSV processing tasks corrections --> CSV generation ("Data rescue" dataset updated) --> import into Clisys --> errors and rejection checks ...

As many time as required to fix all the issues.

This could build a proper "staging dataset", usable in Clisys for advanced quality control.

Once controlled in Clisys (with potential manual data modifications), "staging dataset" is CSV exported and reimported into Clisys production dataset.

Process c: migrate the data from the **Data Rescue Schema (Step 3)** to the **Data Staging Schema (Step 4)**. This will be a very challenging aspect of the Pilot Project and is where both climatological and database management) expertise will be required. This step could be performed by mixed means: SQL language, shell script, program language, application, CDMS and database tools. This will be an iterative process with repeated errors and rejections checks, in which data sets that need correction are reprocessed, put into CSV format again, then re-imported into Clisys (staging) as many times as needed until a proper staging data set is built and can be subjected to advanced quality control within Clisys. Once the datasets have been through this process and have passed through the complete quality control process, data will be exported as CSV formats.

Process d: import the quality controlled CSV formats from the **Data Staging schema 4** to the **Data Production schema 5** where the rescued data is now operationally in place. The Clisys CDMS tool is used for the final import.

Technical observation by MFI: Clisys cannot copy the data from one set of parameters to another. This will be accomplished by exporting the CSV-formatted data in the Data Staging Schema and re-importing them into the Data Production Schema.

Process e: Analyse data completeness and possible needs for further digitization and data rescue - Compare digital inventory and priority data to identify data that have yet to be digitized and prioritize these data vis-a-vis climate modeling requirements at the national and regional levels.

This comparison of the digital observations and imaged observation datasets will be undertaken by querying the data within the CDMS; the CDMS will facilitate identifying the gaps in the digital database and determining which data in paper formats still require digitization.

As part of the extensive Angola Data Rescue Project (project Phases following the Pilot Phase), the newly identified yet-to-be digitized data can be digitized and ingested and operationalized in the CDMS.

Guiding principles

This approach allows a large spectrum of tools to be used and does not recommend specific tools. Nevertheless, the approach relies on the following climatological principles, reusability principles and capacity development principles:

- the data rescue-related work carried-out by/in support of INAMET (whose CDMS is Clisys) can be applied to any other CDMS (CLIDE, CLIMSOFT, CLIWARE, MCH, MICROSTEP, etc.);
- data provenance and traceability will be ensured;
- data access and system security will be ensured; and
- the capacity of INAMET to make use of its CDMS will be enhanced.

The tools to be used, whether such tools are external or internal to the CDMS, will depend on technical guidance provided by the Technical Experts and should reflect the capacity of the INAMET staff and the external technicians that will be involved with/engaged for the project.

Appendix 2
Supplemental information on deliverables and managing system security

Potential detailed deliverables and system security

To maximize the system security, only agents from INAMET with the assistance of MFI shall intervene in processes that modify the operational system in place at INAMET.

The following seven deliverables are considered below:

Deliverable I – Unified inventory

Deliverable II – Table definition

Deliverable III- Preprocess in CSV format

Deliverable IV – Data Rescue Schema integration

Deliverable V - Data Staging Schema Integration

Deliverable VI - Data import into Operational CDMS schema

Deliverable VII - Analysis and comparison

- Deliverables I, II, III could be performed by a non-Clisys administrator;
- Deliverables IV, V and VI shall be performed under the supervision of a Clisys administrator.

Name	Deliverable/sub-deliverable	Step/process	Performed under CDMS administrator supervision
Deliverable I - Unified inventory	Unified inventory	Step 1	No
Deliverable II – Table definition	Table definition	Part of Step 3	No
Deliverable III - Preprocess in CSV format	● Definition of ingestible CSV format for Clisys. ● Creation of usable Scripts for each dataset type (NCEI/ULisbon), which can be used for batched data tables. ● CDMS appropriate CSV files created from selected datasets and stored in accessible Relational Database ● Written procedures for preparing scripts. ● Hire of 2 contractors, preparation of TOR, etc	Process (a)	No Support required to define CSV format required by CDMS.
Deliverable IV - Data Rescue Schema integration	● CDMS appropriate CSV files import into Clisys Staging (non-operational). ● Written procedures for ingesting historic data into CLISYS	Step 3 & Process (b)	Yes

Deliverable V -Data Staging Schema Integration	● Final output: CSV files comprising the corrected data staging set, reviewed and controlled by IT/Climatologists. ● Datasets reviewed and corrected as necessary, with production of corrected CSV files, and re-import as necessary to correct errors to achieve the final output. ● Written procedures for this quality control process.	Step 4 Process (c)	Yes
Deliverable VI - Data import into Operational CDMS schema	● Corrected and reviewed CSV files prepared and vetted for the operational system are imported into Data Production Schema.	Process (d)	Yes
Deliverable VII - Analysis and comparison	● Identification of imaged datasets that have not yet been digitized through inventory comparison of digital datasets and imaged datasets. ● Final Report, to include all procedures, examples of scripts, CSV (pre-process, final), and analysis and comparison.	Process (e)	Yes

For each deliverable, the following figures will be defined:

- A person/organization responsible of the deliverable;
- A person/organization in charge of the operations/processes needed for the deliverable;
- A person/organization in charge of the verification;
- A person/organization in charge of the writing of the acceptance tests.

Deliverable I – Unified inventory

Unified Inventory for all known sources, inventory should be made available globally, regionally and nationally. Selection of the data from NCEI and from the University of Lisbon.

e.g.: Globally with the Data Rescue Portal, Regionally with SADC, Nationally with a shared drive between the actors.

Identify the figure that is:

- in charge of the deliverable -
- in charge of the operations/processes
- in charge of the verification -
- writing the acceptance tests -

Deliverable II -Table definition

Development of Relational Database Management System tables description that will ingest the data from NCEI and University of Lisbon selected for Deliverable 1.. Such description should be enough to deploy the tables in an Oracle environment and in other databases environment (MySQL, POSTGRESQL, etc.).

Such development will be available to WMO to be shared with NMHSs.

Identify the figure that is:

- in charge of the deliverable -
- in charge of the operations/processes -
- in charge of the verification -
- writing the acceptance tests -

Deliverable III -Preprocess in CSV format

Acquisition of the data from NCEI and University of Lisbon selected from Deliverable 1

Development of scripts that will allow to produce csv file in accordance with the tables defined by deliverable II. It is preferable that such scripts could be run in multi-environment operating system (e.g. Linux, Windows)

Make this development open source and share it with WMO, INAMET, SADC, University of Lisbon.

A document that details all the processes developed for WMO and INAMET.

Identify the figure that is:

- in charge of the deliverable -
- in charge of the operations/processes -
- in charge of the verification -
- writing the acceptance tests -

Deliverable IV – Data Rescue Schema integration

Integrating the work of deliverable II and III into the INAMET CDMS.

A document that details all the processes developed.

Identify the figure that is:

- in charge of the deliverable -
- in charge of the operations/processes -
- in charge of the verification –
- writing the acceptance tests -

Deliverable V – Data Staging Schema integration

Development of the Data Staging schema for the data from NCEI and University of Lisbon selected for Deliverable 1. And technical document to allow specific data to be ingested into the Data Staging schema.

Document that details all the processes developed

Identify the figure that is:

- in charge of the deliverable -
- in charge of the operations/processes -
- in charge of the verification –
- writing the acceptance tests -

Deliverable VI – Data Staging Schema–quality control, data assessment process

Use Clisys to review datasets and correct errors, repeated as necessary, with the final output of a FINAL, quality-controlled CSV format that can then be ingested into the CliSys Production Schema.

Document that details all the processes developed.

Identify the figure that is:

- in charge of the deliverable – INAMET IT and climatology/MFI
- in charge of the operations/processes -
- in charge of the verification -
- writing the acceptance tests -

Deliverable VII – Import data into Data Production schema

CDMS import into the Data Production schema.

Document that details all the processes developed and that gives an analysis of the data imported: number of stations, parameters, % missing data, etc.

Identify the figure that is:

- in charge of the deliverable – INAMET Database Manager?
- in charge of the operations/processes -
- in charge of the verification –
- writing the acceptance tests -