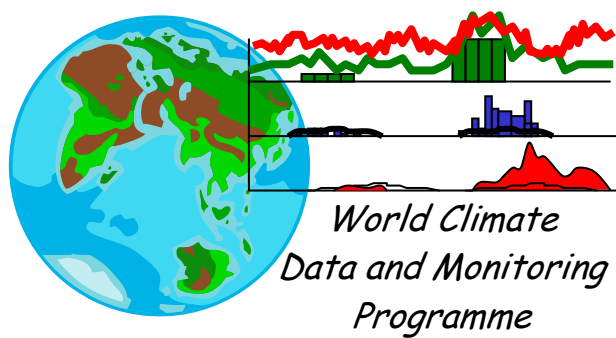


GUIDELINES ON CLIMATE DATA MANAGEMENT

WCDMP-No. 60

WMO-TD No. 1376



World Meteorological Organization

(Geneva, March 2007)

(Contributions by: N. Plummer, W. Lipa, S. Palmer, G. Prank, J. Shortridge,
D. Stuber)

Edited by: Omar Baddour and Hama Kontongomde

©2007 World Meteorological Organization

WMO/TD No. 1376

NOTE

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of any of the participating agencies concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

This document is not an official WMO publication and has not been subjected to WMO's standard editorial procedures. The views expressed herein do not necessarily have the endorsement of the WMO.

THE WCDMP “GUIDELINES” SERIES

In recognizing the need for National Meteorological Services (NMHSs) to improve their climate data and monitoring services, the WMO' Commission for Climatology (CCI) placed a high priority on the distribution of guidelines for the NMHSs.

Within the World Climate Data and Monitoring Programme, under the Open Programme Area Group (OPAG I) on Climate Data and Data Management the Expert Team on Data Management initiated the preparation of this guidelines Document. These guidelines are intended to provide NMHSs with information on best practices in climate data management and assist them in making the transition from older databases, such as CLICOM, to the kind of systems that are providing much greater utility, security and robustness.

The Guidelines document was drafted by a sub-group of the CCI Expert Team on Climate Data Management and reviewed externally. During its first meeting in Nairobi, 1-3 November 2006, the newly appointed Expert Team by the fourteenth CCI session (Beijing, China , 3-10 November 2005), proposed to make a second revision of the document and provided few updates.

It should be kept in mind that this Technical Document, like the other technical documents published under the WMO WCDMP series, is intended to provide guidance in the form of best practices that can be used by Members. Because of the diversity of NMHSs, with respect to size and stage of technological development, it may not have a significant utility for specific Members. However, this document does cover a wide range of guidance that should provide some form of assistance to every Member.

Guidelines on Climate Data Management

Neil Plummer¹, Wolfgang Lipa², Steve Palmer³, Glenn Prank⁴,
John Shortridge⁵, Denis Stuber⁶

Contents

1	INTRODUCTION AND PURPOSE	7
2	BRIEF HISTORY OF CLIMATE DATA MANAGEMENT	8
3	CLIMATE DATA MANAGEMENT AND ORGANIZATIONAL CONTEXT	9
3.1	User requirements and supporting priority needs	9
3.2	Climate Data Management Systems: Desirable properties	10
	3.2.1 Database model	10
	3.2.2 Key entry capabilities	11
	3.2.3 Electronic input options	11
	3.2.4 Quality issues and associated metadata	11
	3.2.5 Scope of quality checks on observation values	12
	3.2.6 Data extraction	12
3.3	Security issues	12
3.4	Database management and monitoring	13
3.5	Documentation management	14
4	ESSENTIALS OF CLIMATE DATA FLOW MANAGEMENT	15
4.1	Metadata documentation and management	15
4.2	Data acquisition, entry, storage and archiving	17
	4.2.1 Acquisition	17
	4.2.2 Entry	17
	4.2.3 Storage and archiving of hard copy records	18
	4.2.4 Storage and archiving of digital information	19
4.3	Managing original records and data rescue	19
4.4	Quality assurance and quality control	21
4.5	Data exchange	23
4.6	Data access and product development	24
4.7	Data administration and monitoring	25
4.8	Change management issues	26

¹ Bureau of Meteorology, Melbourne, Australia

² Central Institute for Meteorology and Geodynamics (ZAMG), Vienna, Austria

³ Met Office, Exeter, UK

⁴ Bureau of Meteorology, Melbourne, Australia

⁵ Bureau of Meteorology, Melbourne, Australia

⁶ Météo France, Toulouse, France

5	TRANSITION BETWEEN A DATABASE MANAGEMENT SYSTEM	27
5.1	Choosing a climate database management system: what needs to be considered?	28
5.1.1	Consider the business needs	28
5.1.2	Designing the required system	28
5.1.3	Scalability	29
5.1.4	Architecture and technology	29
5.1.5	Choice of CDMS	30
5.2	Database architecture considerations	31
5.2.1	Database design considerations: normalization	32
5.2.2	Data models in use by CDMSs	32
5.3	Computer hardware and software considerations	34
5.3.1	Current situation analysis	35
5.3.2	Functional solution design: size up the requirements	35
5.3.3	The technical solution design	38
5.3.4	System and service transition issues	42
5.4	Making the transition from CLICOM	43
5.4.1	Level of expertise required	43
5.4.2	Secure the existing CLICOM system	43
5.4.3	Prepare the metadata to be imported	43
5.4.4	Prepare the climate data to be imported	44
5.4.5	Collect information on CLICOM quality control set up	45
5.4.6	Intensive testing of the new system	45
6	SUSTAINING DATA MANAGEMENT OPERATIONS	46
6.1	Resource requirements, including staffing	46
6.2	Training	46
6.3	Occupational health and safety issues	47
7	CONCLUSIONS	47
8	ACKNOWLEDGMENTS	48
9	REFERENCES	48
10	GLOSSARY AND DEFINITIONS	49
	APPENDIX 1 CDMS DESCRIPTIONS	55
	APPENDIX 2 CASE STUDY OF CDMS IMPLEMENTATION	57
	APPENDIX 3 HARDWARE EXAMPLES	61

Introduction and purpose

The World Climate Data and Monitoring Programme (WCDMP)/Commission for Climatology (CCI) series of guideline documents provide information and assistance on how to organize and implement climate services in areas of importance to a National Meteorological and Hydrological Service (NMHS). They are also designed to present processes and technological solutions that attempt to address the special situation and needs of small NMHSs, which often have limited resources. To the extent possible, guideline documents should also be:

- Applicable to all other WMO programmes;
- Flexible to a range of needs;
- Formulated in a way that can be upgraded; and
- In accordance with the principles of, and act as a complement to, the Guide to Climatological Practices.

These guidelines are intended to provide NMHSs with information on best practice climate data management. This information is timely given the need for many countries to make the transition from older databases, such as CLICOM, to the kind of systems that are providing much greater utility, security and robustness.

Given the overall objectives of these guideline documents, some emphasis will be given to building knowledge and capacity in:

- The phases and stages of managing computerized data;
- Available database technologies and selecting an appropriate database;
- Making the transition to a modern climate database system; and
- Sustaining data management operations.

There is a deliberate focus on those aspects of climate data management that are of interest to NMHSs wishing to make the transition to a modern climate database management system and, just as important, on what skills, systems and processes need to be in place to ensure that operations are sustained. These Guidelines are not directed at NMHSs wishing to build a database 'from the ground-up' but are more appropriate for those wanting to analyze and evaluate candidate systems.

The details of data management practices and standards are omitted from these guidelines as these are generally documented in the World Meteorological Organization (WMO) *Guide to Climatological Practices* (WMO 2005) and in other guideline documents (e.g. WMO 2003a, 2003b, 2004). Readers wishing to learn more about the technical issues of databases are referred to specific texts (e.g. Date 1999). The WCDMP series of publications (at <http://web/wcp/wcdmp/html/wcdmpreplist.html>) are also an extremely valuable resource.

While an in-depth knowledge of climate data management is not assumed, readers who have a working knowledge of this topic will be more likely to benefit from these guidelines.

Section 2 will briefly discuss the historical development of climate data management with respect to NMHSs. Section 3 will discuss climate data management and organization context, i.e. those broader level issues that data managers need to be aware of. More detailed discussion of the topic is reserved for Section 4 where managing data flows is covered. The

transition between database management systems follows in Section 5, including issues involved in the transition from CLICOM to a modern system. The key issues involved in sustaining data management operations is in Section 6 and we end with some Conclusions in Section 7.

Brief history of climate data management

Information about the weather has been recorded in manuscript form for many centuries. The early records included notes on extreme and, sometimes, catastrophic events and also on phenomena such as the freezing and thawing dates of rivers, lakes and seas, which have taken on a higher profile with recent concerns about climate change.

Specific journals for the collection and retention of climatological information have been used over the last two or three centuries (WMO 2005). The development of instrumentation to quantify meteorological phenomenon and the dedication of observers in maintaining methodical, reliable and well-documented records paved the way for the organized management of climate data. Since the 1940s, standardized forms and procedures gradually became more prevalent and, once computer systems were being used by NMHSs, these forms greatly assisted the computerized data entry process and consequently the development of computer data archives.

The latter part of the twentieth century saw the routine exchange of weather data in digital form and many meteorological and related data centres took the opportunity to directly capture and store these in their databases. Much was learnt about automatic methods of collecting and processing meteorological data in the late 1950s, a period which included the International Geophysical Year and the establishment of the World Weather Watch. The WMO's development of international guidelines and standards for climate data management and data exchange assisted NMHSs in organizing their data management activities and, less directly, also furthered the development of regional and global databases. Today, the management of climate records requires a systematic approach that encompasses paper records, microfilm/microfiche records and digital records, where the latter include image files as well as the traditional alphanumeric representation.

Prior to electronic computers, mechanical devices played an important part in the development of data management. Calculations were made using comptometers, for example, with the results being recorded on paper. A major advance occurred with the introduction of the Hollerith system of punch-cards, sorters and tabulators. These cards, with a series of punched holes recording the values of the meteorological variables, were passed through the sorting and tabulating machines enabling more efficient calculation of statistics.

The 1960s and 1970s saw several NMHSs implementing electronic computers and gradually the information from many millions of punched cards was transferred to magnetic tape. These computers were replaced with increasingly powerful mainframe systems and data were made available on-line through developments in disk technology.

A major step forward occurred with the WCDMP's CLICOM project. Initiated in 1985, this project led to the installation of PC-based climate database software, backed up with hardware and a comprehensive training programme, in more than 100 NMHSs around the world. The project also provided the foundations for demonstrable improvements in climate services, applications and research in these countries. In the late 1990s, the WCDMP initiated a Climate Database

Management System (CDMS) project to take advantage of the latest technologies to meet the varied and growing data management needs of members. Although sometimes more expensive and complex from a technological perspective (and therefore requiring more specialist skills to administer), the new CDMSs offer improved data access and security and much greater utility for users. Relational databases had proven their worth in several NMHSs.

Aside from advances in database technologies, more efficient data capture was made possible through the mid-to-late 1990s with an increase in automatic weather stations (AWSs), electronic field books (i.e. on-station notebook computers used to enter, quality control and transmit observations), the Internet and other advances in technology.

Not surprisingly, there are a number of trends already underway that suggest there are many further benefits for NMHSs in managing data and servicing their clients. The Internet is already delivering greatly improved data access capabilities and, providing security issues are managed, we can expect major opportunities for data managers in the next five to ten years. In addition, Open Source⁷ relational database systems may also remove the cost barriers to relational databases for many NMHSs over this period.

Climate data management and organizational context

User requirements and supporting priority needs

It is essential that both the development of climate databases and the implementation of data management practices take into account the needs of the existing, and to the extent that it is predictable, future data users. While at first sight this may seem intuitive, it is not difficult to envisage situations where, for example, data structures have been developed that omit data important for a useful application or where a data centre commits too little of its resources to checking the quality of data for which users demand high quality. Where practicable it is suggested that all measurements taken are stored in CDMS even where there is no current use made of this data.

In all new developments, data managers should either attempt to have at least one key data user as part of their project team or undertake some regular consultative process with a group of user stakeholders. Data providers or data users within the organization may also have consultative processes with end users of climate data (or information) and data managers should endeavour to keep abreast of both changes in needs and any issues that user communities have. Put simply, data management requires awareness of the needs of the end users.

At present, the key demand factors for data managers are coming from climate prediction, climate change, agriculture and other primary industries, health, disaster/emergency management, energy, natural resource management (including water), sustainability, urban planning and design, finance and insurance. Data managers must remain cognizant that the existence of the data management operation is contingent on the centre delivering social, economic and environmental benefit to the user communities it serves. It is important, therefore,

⁷ Open Source refers to computer software whose source code is freely available for others to view and modify. This means that the software is available free of charge (or sometimes for the cost of distribution).

for the data manager to encourage and, to the extent possible, collaborate in projects which demonstrate the value of its data resource. Even an awareness of studies that show, for example, the economic benefits from climate predictions or the social benefits from having climate data used in a health warning system, can be useful in reminding senior NMHS managers or convincing funding agencies that data are worth investing in. Increasingly, value is being delivered through integrating data with application models (e.g. crop simulation models, economic models) and so integration issues should be considered in the design of new data structures.

Effectively satisfying user needs will rely on the appropriate organizational structures and responsibilities being in place. For example, the organization needs to be open to external opinions and also have effective and efficient communication pathways that ensure information is interchanged internally.

Data managers also need to be mindful of the principles for long-term climate monitoring as stated in GCOS (2003), endorsed by the Commission for Climatology (CCI) and adopted by the United Nations Framework Convention on Climate Change. The ability to maintain the continuity, homogeneity and, ultimately, quality of climate data is greatly influenced by how well observation networks and systems are managed. In the context of the ten climate monitoring principles (refer WMO (2003a) for a discussion on how these apply to the operations of an NMHS) it is critical that data managers build close relationships with observation managers as well as climate data users. The data manager will need to operate as an effective intermediary between the observation manager and data user in presenting the former with requirements and, for the latter, in balancing needs with economic and technical feasibility. In the context of climate observation networks and systems, WMO (2003a) explores these principles further.

In responding to user needs, some of the key issues to consider in prioritizing new/additional observations are:

- National social, economic and environmental priorities;
- Data-poor regions;
- Poorly-observed parameters;
- Regions sensitive to change; and
- Measurements with inadequate temporal resolution.

Climate Data Management Systems: Desirable properties

3.2.1 Database model

Any climate database will be based on some underlying model of the data. This model is very important for the quality of the resulting system, and in particular for its maintainability (bearing in mind the rough rule that well over 50% of the effort consumed by any software system is the effort consumed by system maintenance and enhancement over the years). An inappropriate model will tend to make the system harder to maintain. In general, a database designed primarily for near-real-time meteorological work will mainly be accessed 'synoptically', since retrievals will typically be of the form "get all data for some given locations or area and for some defined and relatively short period". By contrast, climate applications typically involve retrieving data for one or a few stations, but for a long period. Consequently, one broad approach to storing climate data involves storing (for daily data) all data associated with a given station and day together. A similar approach can be taken for hourly and monthly data. Another approach

involves storing each different type of datum separately, so that one table (in the context of relational databases) will contain all the air temperature data for all stations and dates, and so on. These issues are given further consideration in Section 5.2. As well as the innate quality of the design, another factor is how well the data model is documented, and how easy it is for programmers to extend it. One key requirement of a data model is that it should be able to represent data conveyed by all the standard WMO message types – TEMP, PILOT, METAR/SPECI, CLIMAT, CLIMAT_TEMP, and SYNOP.

Similar considerations apply to the ‘metadata model’. In a sense this is a more complex area, because the metadata structures are potentially more complex than the data structures. In due course a standard metadata model⁸ should be defined for climate data (extending the work that has already been done on the WMO Core Metadata Standard), and at that time the ability to support this standard model will become a desirable characteristic of climate systems.

3.2.2 Key entry capabilities

The fundamental requirement of a data entry system is that it should not introduce any avoidable inefficiencies into the data capture process. Frequently, staff who undertake this task are highly skilled in the data entry process, so some effort is justified to ensure that they can work productively at all times. This means that the data entry system should be free of annoying defects which may slow up the data entry operator. Ideally the forms presented on the screen will be customizable so as to optimize the efficiency of data entry. The system should also, as far as possible, validate data as they are entered - catching errors and, where possible, suggesting alternative values. It is also possible to present defaults for some items, thus saving unnecessary key strokes. Another useful feature is support for ‘double keying’ as a way of reducing the incidence of errors (where each datum is entered twice, with the data entered by the second operator being automatically compared with what was entered by the first operator).

3.2.3 Electronic input options

As mentioned above, it is desirable for a CDMS to have the ability to represent the full content of the relevant WMO standard message formats – SYNOP, CLIMAT, etc). An associated beneficial feature is the ability to decode these message formats directly into the climate database. In some cases (notably the monthly CLIMAT and CLIMAT TEMP messages) it is also desirable to have systems which will generate these messages from information held in the database. Other useful electronic input options are the ability to input data from AWSs and the ability to ingest data exported from the CLICOM system. In general, this last option would be expected to be used on a once-off basis, as part of the process of system changeover (see Section 5.4).

3.2.4 Quality issues and associated metadata

There is a range of quality information that should be associated with a meteorological datum. This includes the quality ascribed to the datum (which may include such values as “apparently OK”, “suspect”, “inconsistent with some other datum”, etc), the nature of the test(s) undertaken to generate the quality flag, etc. There is also an associated but wider range of relevant metadata. This includes such things as the instrument that was used to make the observation, the ‘audit trail’ of the observation showing all changes made to it since initial data capture, a full

⁸ In this context, metadata refers to information on the climatic data (i.e. data quality, method of derivation, etc) rather than historical information for a station.

range of information about the site, the observations programme in effect at the time, and so on. On a longer timescale, the system will ideally be able to represent the full station history.

3.2.5 Scope of quality checks on observation values

Checks applied to determine the quality of an observation can range from the very simple to the very complex. In roughly increasing order of complexity they can include:

- Syntactic checks (e.g. an air temperature must be a number to at most 1 decimal place);
- Numeric ranges (e.g. the temperature must fall in the range -90 to +70);
- Climate range checks (i.e. is the datum consistent with climatology?);
- Intra-record consistency (e.g. the air temperature must not be less than the dew point);
- Time-series consistency (e.g. the difference between two successive temperatures at a site must be 'plausible'); and
- Spatial consistency (e.g. the station-dependent limits of plausible difference between the temperatures at a station and its neighbours must not be violated).

3.2.6 Data extraction

An important aspect of any CDMS is the range and expressive power of the facilities offered for data retrieval and analysis. Ideally data can be retrieved both from a GUI interface and from a command-line interface, as appropriate. Ideally, GUI retrieval facilities should be provided for the vast majority of users, with query language facilities used only by a small range of knowledgeable users who have a need to do non-standard retrievals. Users who have the ability to do so should be able to specify their own retrieval criteria (through the GUI interface in general), and the system documentation should give as much information as is necessary to support this.

The system should also support a wide range of output options where the broad structure of the retrieval and the presentation of the output are given, and where specifying a retrieval involves customizing for stations, times and details of output presentation. Such options should be provided to give access to listings of data, tabular summaries, statistical analyses (simple and complex), and graphical presentations.

Security issues

The main goals of a security policy and associated activities are to prevent loss of, or damage to, the CDMS and to keep data management facilities in the best possible condition. To satisfy this, the requisites are that:

- The CDMS is in a protected and sheltered building;
- All staff must be aware of their professional responsibility and duty of care;
- The archives and database environment must be secured and protected against fire, humidity, etc;
- Only protected applications permitted to a small group of people have the right to handle data manipulations (i.e. insert, update, delete). People, who have write access to the database must agree not to make any transactions beside the operations and practices

- approved by the data manager. All changes to data tables should have an audit trail and the controls on access to this trail should be in place;
- There should be a policy of not sharing passwords as well as not writing down passwords anywhere. Passwords should be changed regularly and this applies to all users from the database administrator to the user who handles data manipulation applications. The password can consist of seemingly non-related letters or numbers but they could be the leading letters of the words of a whole sentence. For example: I am sitting in the 9th floor at work -> lasit9faw.;
 - The archive database system must run behind a firewall at the highest possible security level;
 - All unnecessary services, like email from the database computer systems should be disabled. The database computer system should only run the kernel of the operating system and the database management system;
 - The database – client computer must also be protected against virus and hacker attacks;
 - Backups must be made at such intervals as to fit with the maximum acceptable loss of work should a failure occur. Typically this would be daily incremental and weekly full backups;
 - Less frequent (typically monthly) archives of the contents of the data tables should be made in ASCII format and put in a safe, secure and fire-proof location, remote from the physical location of the climate database.
 - A recovery plan should be drawn up, setting out how the backups and archives will be used in emergency situations to recover the database. The plan should cover as many disaster scenarios as can be envisaged.
 - The recovery plan should be reviewed and tested at regular intervals.
 - Backups of the CDMS must be performed prior to any changes to system software, or to the design, or to applications contained in the CDMS; and
 - At appropriate intervals, an audit should be carried out of the original document paper archives, and the ASCII format archives.

Database management and monitoring

The aim of database management is to ensure the integrity of the database at all times, and to ensure that the database contains all the data and metadata needed to deliver the objectives of the organization, both now and into the future. Monitoring should use performance indicators to show how well the database and the processes using the database are performing against those objectives.

There will be a number of processes (see Section 3.5) which use the data, including metadata maintenance, supply to the database, QC actions modifying the database, and extraction for end users and customers. Each of these Processes should have a Process Owner, who will be responsible for the Process including the monitoring, evaluation and improvement of the Process. Of course, for a small NMHS, all of these tasks may be part of one Process, and for a very small NMHS, there could be only one person performing most of the roles. It is still useful to understand these different roles and responsibilities, and to separate the actions. As a principle of risk management, when conducting any operation on the database, the user should, as far as possible, have the privileges required to execute that operation, and nothing more.

Typical monitoring reports would include the number and type of stations in the database, the quantity of data in the database grouped by stations and by observation element types. It is

useful to compare these with the observation schedule, to identify where missing data is occurring – action to identify causes and solutions can then follow. Other reports could look at the QC actions, to check that QC is keeping up with the flow of new data, and to identify any groupings of data with excessive QC amendments; again action should follow to rectify problems.

It is very useful to keep track of the quantity and range of data being extracted for end-user enquiries, since this is helpful both in indicating which are the most important sets of data, in justifying the investment in the climate data, and in indicating areas to be developed in future. Where cost-recovery is implemented, it is essential to know the revenue generated by services using the climate data.

The monitoring should measure appropriate Objectively Verifiable Indicators (OVI), and each OVI should have a Means Of Verification (MOV). The MOV provides a target value for the OVI which measures the performance against the business needs. For example, an MOV would be that 90% of monthly reports expected from raingauge-only stations submitted as paper forms should be available in the database by the 15th day of the following month. Note that this would be a SMART target – Specific, Measurable, Achievable, Realistic and Time-bound. The corresponding OVI would therefore compare the number of such reports in the database against the expected value on the 15th of the month, expressed as a percentage.

The frequency and reporting period of monitoring reports will depend on the business needs of the organization. For ingestion of data, some may well be done automatically every day – for this a daily success/failure report is vital to allow remedial action to be taken before the next ingestion cycle. Much climate data has a calendar month cycle, so it would be appropriate to report on the quantity and quality of data ingestion each month. For cost-recovery and cost-benefit analysis, there will need to be reports at least every financial year on the number of end-user products, the income generated from these, and the quantity and type of data involved. The information for this is not necessarily a function of a CDMS, but the CDMS could include usage statistics.

Monitoring reports must be used to actively manage the flow of information, metadata and data through the climate data system of the organization.

In the management planning, consideration also needs to be given to the long-term life of the climate data, so it remains available as a resource for future users. Issues include succession planning and training of personnel, and the replacement cycle for hardware, as well as maintenance and upgrade costs, support and training for commercial software.

Documentation management

Documentation of the processes involved in managing and using the database is essential, both to record the design and as operational instructions and guidelines for the managers, users and developers of the database and associated applications. For those organizations which have implemented an ISO9000 Quality Management System, this documentation is mandatory for compliance. For other organizations, documentation using the ISO9000 principles is highly recommended.

The ISO9000 principles can be summarized as:

- Say what you do;
- Do what you say;
- Check and monitor that you are doing what you say; and
- Evaluate and improve.

The documentation should be written in a positive way to encourage best practice by the people operating the process.

Management and documentation should use the concept of Processes, where a Process has inputs; a series of operations are performed on the inputs; and outputs are produced for a 'customer'. The 'customer' may be internal or external to the organization. For example, the 'key-entry process' could include booking-in of paper forms, delivery of those to key-entry operators, the keying of the data into the database form, loading of the completed form into the database and resolution of any errors, and delivery of the paper forms to the archives. Inputs are the paper forms, the staff and computing resources. Outputs would be data in the database and paper forms in the archive. The customer would be users of the database, including other Processes such as Quality Control.

Each Process should have an Owner, who is responsible for the documentation and overall management and monitoring of the Process. Changes to the Process itself, and to the documentation need to be agreed and authorized by the Process Owner. Documentation which is out-of-date or wrong could lead to destruction of the database. Similarly, a failure to carry out the documented actions can also lead to loss of the database.

The Process Owner should also lead and encourage the evaluation and improvement of the Process, to ensure that it remains fit for purpose and meets the objectives of the organization.

Essentials of climate data flow management

Metadata documentation and management

In order that meteorological data be useful for future users, it is essential that an adequate set of metadata be available. Such metadata records (ideally) state precisely where the measurements were made, by whom, using what instruments, the quality that has been ascribed to the data, etc. A detailed treatment of station specific metadata can be found in the report "Guidelines on Climate Metadata and Homogenization" (WMO 2003b) in the World Climate Data and Monitoring Programme series of publications. The reader is referred to that publication for an overall treatment of the issue, while this document attempts to give a quick overview of systems-related issues.

An important general point in this context is that the innate structure of the metadata in an ideal system will be considerably more complex than the innate structure of the climate data itself. Consider a rainfall observation. Commonly, the essential information content of the data will be simply of the general form "at station x, for an elapsed period of time e, ending at time t, the precipitation was p mm." The associated metadata, which is needed to fully interpret the data, could include such things as the:

- Reference date used by the database (GMT, time zone, others);

- Quality which has been ascribed to the observation;
- History of the values ascribed to the meteorological parameter and any associated flags;
- Instrument used to record the observation, together with more fine detail on its own maintenance programme, tolerances, internal parameters, etc;
- Name of the observer;
- Full details of the station and its history;
- Programme of observations in effect at the time and its history;
- the inventory of the elements stored in the database, their units, their boundaries; and
- Topographical and ground-cover details of the site, information on surrounding trees, buildings, etc, and how these have changed with time.

Recording information at this level of detail in a computer system will require a complex data structure, and correspondingly complex software. It will also require the ability to store graphics (site plans, photos, possibly images of old documents, etc). This is not conceptually complex, but has implications for, in particular, the hardware to be used in order to capture and reproduce a range of images, and the underlying software to be used as the database management system.

There is also the need to provide metadata at the dataset level but details of this is beyond the scope of these guidelines.

Data may also be the subject of homogenization work, which is intended to remove from a data set all influences other than those from the underlying climate. Data that has been the subject of such exercises has a need for additional associated metadata, in order to accurately represent the nature of the homogenization work. Homogenized data sets should be a product of the CDMS and never considered as a replacement.

Sophisticated tools for managing and accessing metadata are becoming increasingly available (Figure 1).

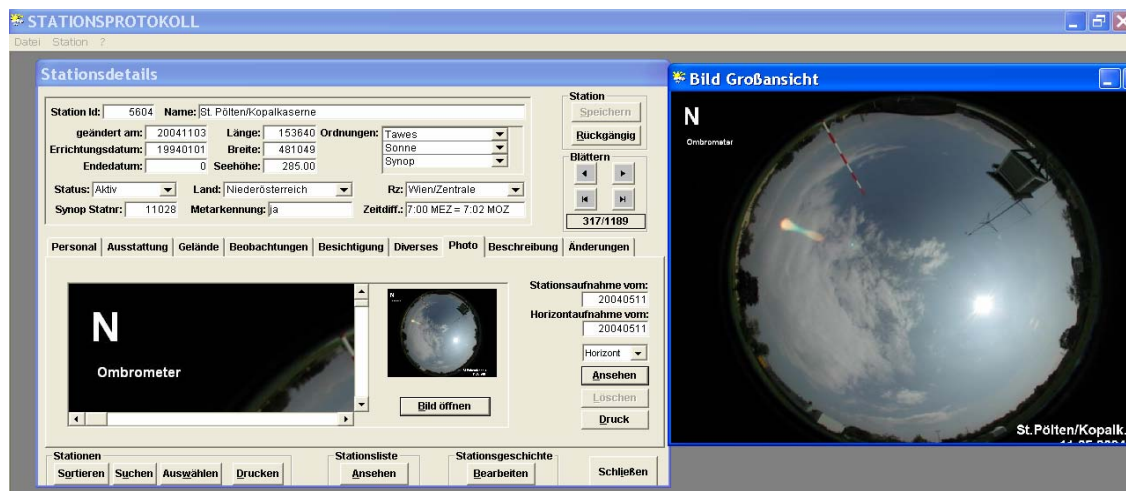


Figure 1. The ZAMG (Central Institute for Meteorology and Geodynamics, Austria) Metadata Management Software STATPROT.

Data acquisition, entry, storage and archiving

4.2.1 Acquisition

One of the main tasks of a NMHS is to run a climate observation network as described in WMO (2003a). While providing the 'life blood' for an NMHS, these networks are run at a cost to the organization and generally provide little immediate commercial benefit. The overheads associated with running a network could be reduced if the NMHS were to use the station networks of similar institutions such as universities, local utilities, and private companies. In these cases all relevant climate data should be collected and it is highly desirable that the NMHS should be granted full use of all the climate data, without restriction, as if it were its own data. Appropriate contractual, or 'Memorandum of Understandings', between the NMHS and other organisations may need to be drafted and signed at senior management level.

Non-conventional data and information can be managed. For example, there exists a lot of governmental and privately owned web-cameras and, if there is a business need, you can capture and manage this information. Other useful information sources could be gained from the print media, radio or television. If resources permit, all reports about weather-related disasters and other interesting weather news could be collected and put in a document management system, which can be a part of a broader climate database.

4.2.2 Entry

Data collection should be as close to the source as possible. AWSs, including stations which have some manual observations, and partially automatic weather stations should collect their climate data and error messages on site and transfer these electronically to the CDMS, possibly via another database system. Manually observed data should be collected and captured on-site and transferred as soon as possible to the CDMS. In the case of a manual station the data should be collected and captured on-site and a diurnal digital transfer to the CDMS is highly desirable. While it is noted that it may not always be possible to get all data stored in the database within a day (e.g. monthly returns of paper records), the advantages of a daily database collection are that:

- Data quality is likely to be improved. For example, it is easier for a quality control operator to remember yesterday's weather, but remembering the weather 30 days before is far more difficult;
- While the cost for data transmission will increase, the human effort for quality control will likely decrease and there will be greater opportunities from improved access to more data; and
- Technical errors will be detected much faster.

Of course many stations only record their data on paper and a monthly, or a within-month, transmission through the postal service will be required. A record of incoming data reports should be maintained on paper or spreadsheet or, ideally, be generated automatically and compared with what data were expected. The latter allows missing reports to be easily detected

and also allows for follow-up with the observer. Before entering the data from the autographic reports, they must be evaluated.

For data capture, a system that checks the constraints and data type for each parameter before it ingests the values into the database is recommended.

Data entry and data quality control are strictly related since some QC is often performed during data entry. However, in general, the digital climate data still needs to go through a subsequent QC process prior to it being classified as being fully quality controlled. End users should always be made aware of the QC status of data they are provided.

It is important to retain the originally received data value (for both evaluation of QC processes and for legal and similar enquiries – see Section 4.3) as well as the latest quality controlled value. The original received value will only have passed through the type and range checking of the database ingestion process and, even if rejected by this process, it still needs to be retained. The latest quality controlled value may change as the stages of QC are performed.

The use of non-traditional sources and types of data should be allowed for, depending on the operating practices of the organisation, e.g. the use of web-cams for remote observing.

The CDMS could contain information about media reports, pictures, etc, beyond the traditional data and metadata. Capture of such information could be through:

- Obtaining an image of the report from the print media with a digital camera or scanner;
- Defining date, region and type of action (flood, drought, heavy precipitation) and name of the media;
- Giving some comments about the event; and
- Storing the information in an archive system.

Data capture of extreme weather, or other noteworthy events, from television or radio broadcasters can be captured in a similar manner.

4.2.3 Storage and archiving of hard copy records

All paper records should be stored in a controlled environment to avoid deterioration and possible destruction by temperature and humidity extremes, insects, pests, fire, flood, accidents or vandalism. But before archiving, the records should be captured in microfilm or, preferably, in electronic image form through a digital camera or scanner as is described in WMO (2004). It is recommended to record metadata on the microfilm and/or image archive, which can then become part of a broader climate database.

The advantage of records in electronic image form is that:

- You receive an additional 'digital' copy for security reasons;
- A digital copy of a digital copy is easy to make and can be stored for security reasons in different places;
- If you need access to the original data, you can use the electronic image copy and prevent the original paper from deterioration; and
- Access to the records can be provided through the Internet or by email and thus satisfying a greater number of users.

Evaluated autographic reports should also be treated like the paper report.

The paper archive should be administrated in such a way, that

- Old data (e.g. more than one year old) should be stored in chronological order; and
- The remainder in order of synoptic data elements.

4.2.4 Storage and archiving of digital information

Here we distinguish data storage, which may be a transitory phase for the data as it moves from collection to an archive form and where the latter generally means that some form of value-adding has taken place (e.g. QC, conversion into different measurement units).

The selection of data to store and archive can sometimes be difficult. AWSs will often generate data that are relevant to the quality of observations but which are not strictly climate data (e.g. information on the battery-level voltage for an AWS). As a general rule, this information should be utilized (e.g. in the quality control process) prior to data archiving and the information not included in the CDMS. (All important metadata should be though, see Sections 3.2.4, 4.1 and also WMO (2003b)). Again, the original message should be stored permanently and made accessible to data managers.

An important job of the data manager is to estimate data storage requirements, including estimating future growth. Account must be taken of the additional information to be included in data records (e.g. QC flags, original messages, date/time of record update), metadata needs and any redundancy necessary to ensure that databases can be restored. Estimating future growth will be difficult but it is worth keeping in mind that many NMHSs are now archiving 1 and/or 10 minute AWS data and that non-traditional climate data (e.g. soil moisture, vegetation indices) are very important for climatologists. The storage requirements for remotely sensed and oceanographic data are generally very large.

In the case of having not enough mass storage to keep all original raw data, the oldest data could be rolled out of the database to a slow mass storage archiving system. This would generally be a tape robot system but nowadays it is often a DVD robot system or something similar. If necessary, a roll-in procedure should reload back small parts of raw data.

If, after some period of time and after the application of QC, there is no requirement to maintain an archive of the original digital data in the CDMS, these can be deleted provided copies of the originals were held elsewhere in the organisation.

Managing original records and data rescue

Data rescue has been comprehensively covered in other WCDMP documents (e.g. WMO 2002, 2004) and so will be dealt with only briefly here.

The issue of managing original records, while related to data rescue, comes with its own set of issues. There are a number of reasons for permanently archiving the first instance of an observation, i.e. the value of a climate parameter that has not been changed, appended to, or removed by either manual or automated processes. Firstly, if users are making decisions based on data in near real-time then there may be legal reasons why an organization should wish to

retain the original value. Secondly, the performance monitoring of observation networks and systems will be biased if the evidence for a poorly performing station is being discarded (e.g. an erroneous data value is being removed).

Ideally, the original data values should be retained in the CDMS and should be as accessible to internal staff as the quality controlled values. Management of these original values will require careful use of quality control flags (see Section 4.4) and the database manager may want to prevent direct external access to these data or, at the very least, take measures to ensure that users are fully aware of the quality of data they have accessed. If it is not possible to store an electronic copy of the original data then hard copy originals should be preserved in some way, consistent with the WMO's recommendations on data rescue.

WMO (2002) defines data rescue as "an ongoing process of preserving all data at risk of being lost due to deterioration of the medium, and the digitization of current and past data into computer compatible form for easy access." This definition implies that:

- Data should be stored as image files onto media that can be regularly renewed to prevent the deterioration of the medium (cartridges, CDs, DVDs etc);
- Data already in computer compatible media should be constantly migrated to storage facilities that conform to changing technologies; and
- Data should be key-entered in a form that can be used for analyses.

Data rescue efforts are essential in order to provide immediate and wide public access to a valuable national resource. From another perspective, data rescue can be considered as preserving the heritage of a country and this fact, along with the many other benefits from climate data, needs to be communicated to senior managers within NMHSs and to others within government in order to secure funding for data rescue activities. WMO (2002) describe some success stories, such as the Belgium funded program which assisted more than 40 African countries in preserving many millions of records on microfiche.

WMO (2004) describes the essentials of data rescue, namely the stages of:

- Search and locate;
- Creating inventories;
- Preservation and storage;
- Validating imaged files;
- Key entry and quality control; and
- Making data accessible.

Figure 2 illustrates these processes.



Figure 2. Processes involved in data rescue.

WMO (2004) also describes the relevant computing technologies, media alternatives and data migration strategies. In terms of capturing electronic images, there is a preference for using digital cameras instead of scanners, although the latter will have advantages in some circumstances (e.g. multi-page scanners can efficiently capture images from multiple paper forms). Data rescue projects should always include a training component, which provide NMHS staff with skills in all facets of data rescue activities and will help ensure that knowledge can be passed on to staff being introduced to data rescue. Occupational health and safety is particularly important for data rescue. For example, staff will be dealing with dusty and heavy paper records in storage facilities that may be inhabited by rodents or other creatures.

Quality assurance and quality control

The quality control process is at the core of the whole data-flow process. It is responsible for ensuring that data is checked and, to the extent possible, is error-free. All errors and mistakes coming from the station site, instrument/sensor, data transmission or data entry stages must be detected and eliminated and, if possible, these should be replaced by correct values (while retaining the original values). A requisite part of the quality management of a NMHS is the data quality control process.

The first step is to build up a logical quality control model, which describes the differences between the checks and tests and flagging. The scope of the quality checks is described in Section 3.2.5.

The flagging model should describe if a value is:

- Checked or not;
- Original or not;
- Suspect or not;
- Erroneous or not;
- Correct or not; and
- Calculated (or derived) or not.

The flagging model can also describe if a value has been:

- Checked automatically or by an operator;
- Calculated with non-suspect values; and
- Calculated with non-missing values.

Data should flow through the complete quality control system and options to bypass this system should be prevented. One or more data flow charts should describe the whole data process, beginning from station network up to the application level with a focus on the quality control process. Manuals and handbooks, as well as instructions on how to handle the whole data process (and particularly the data quality process) should exist within the organisation and their contents be adhered to. Descriptions of changes to the data quality control system are highly desirable in order to help ensure future improvements and as information to interested data users.

Besides the routine data quality control process, an evaluation of all historical quality controlled data should be made from time to time. The evaluation could be made for each parameter and sensor for all stations or for one station for all parameters. It may be triggered by a client wishing to pay a fee for additional QC to be applied prior to receiving their data.

If, after passing through the quality control process, a particular parameter for a station has many suspect values, then something may be wrong with the instrument/sensor and/or data capture process. If the values from one station have more suspect values than from other stations, there is likely to be a problem regarding the general operation of this station. Where there is manual quality control, care should be taken to ensure that procedures are consistent and valid. For example, an operator who replaces all suspect values with estimates (or removes them altogether) may be doing as much damage to a database as someone who flags all suspect values as correct.

In cases where it is suspected that a faulty instrument/sensor and/or operation of the station are causing problems, the QC operator has to act and, where necessary, notify those responsible for the instrument in order that, for example:

- A change of, or adjustment to, an instrument/sensor is made;
- A station inspection can be undertaken;
- Recommendations for a station shift can be made; or
- The observer is contacted to confirm the suspect extreme value(s).

To ensure the integrity of the quality control process, a process owner must be nominated who is responsible for operating the described process. Each aberration and variation must be detected and revised. From time to time an internal and an external audit should take place to guarantee the integrity of the quality control process. Regular training for all QC operators is essential.

Some useful QC tools are now operating in NMHSs (Figure 3).

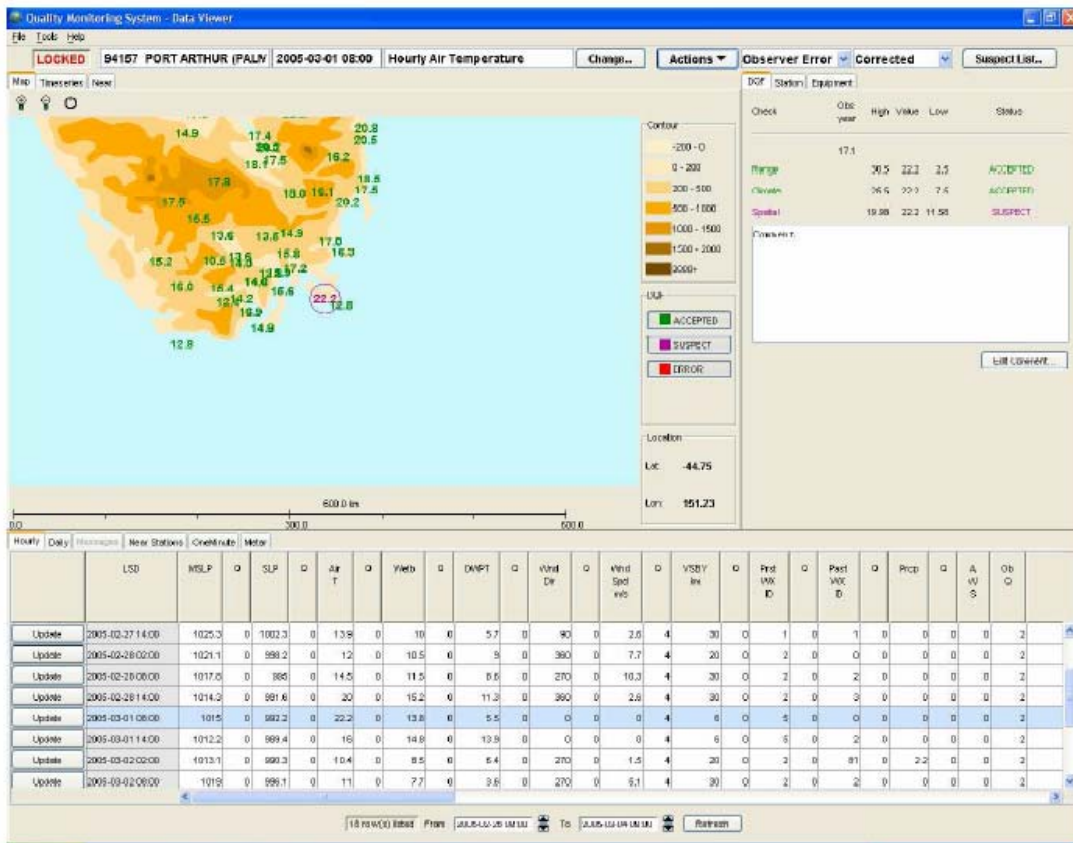


Figure 3. Sample screen shot from a modern Data Quality Control/Monitoring tool – the Australian Bureau of Meteorology’s Quality Monitoring System.

Data exchange

Exchange of data between NMHSs is essential for climatology. This may cover both the storage and use of data (and metadata) from other countries in the database of one NMHS, and the transmission of data to global and regional data centres.

For states which are members of WMO, the obligation to share data and metadata with other members of WMO, and the conditions under which these may be passed to third parties, is covered under WMO Resolution 40 (Cg-XIII) and WMO Resolution 25 (Cg-XIV). These embody the concepts of “essential” and “additional” data, with a specification of a minimum set of data that should be made available with “free and unrestricted access”. Members may decide to declare as “essential” more than this minimum set.

Members of WMO volunteer a subset of their stations to be part of various networks, including the GCOS Upper Air Network (GUAN), the GCOS Surface Network (GSN), the Regional Basic Synoptic Network (RBSN) and Regional Basic Climatological Network (RBCN). Nomination of stations in these networks implies an obligation to share the data internationally. For GUAN, this is mainly through the CLIMAT TEMP message, and for GSN the CLIMAT message. Submission of these messages is monitored, and the results published.

Users are encouraged to study these Resolutions carefully as part of implementing national data policy, which will determine aspects of the CDMS.

There are perhaps four categories of climate data to consider from the data exchange viewpoint:

- Real-time or near-real-time data exchanged over the GTS, and for which timeliness is vital;
- Delayed-mode datasets which may be composites of the above but which are limited in terms of QC;
- Delayed-mode datasets which have had extensive QC and possibly homogeneity analysis applied; and
- Other products (e.g. Satellite imagery, metadata, climate monitoring and prediction products, climate indices, national / regional / global analyses and climatologies, outputs from global climate models).

For the first of these, WMO standard formats have been defined. CLIMAT and CLIMAT TEMP are the 'climate-specific' ones, while others - notably SYNOP, PILOT, TEMP and Marine message formats - are relevant to climate work. Details on these can be found in WMO (1995). For the others, there are formats defined for data interchange (for example, HDF and NetCDF) but these are at a higher level of generality and are not, per se, 'self-describing'.

A key current initiative is the WMO's WIS (WMO Information System) project. This is a very large and complex project which, in very brief summary, is intended to provide "a single coordinated global infrastructure for the collection and sharing of information for all WMO and related international programmes". To facilitate this, standard protocols are being developed for the interchange of data and, crucially, metadata. The significance of this for developers and users of CDMSs is twofold. Firstly, the availability of these facilities will allow users of appropriately enabled CDMSs to obtain data easily via standard methods from other organizations that are functioning as Global Information System Centres (GISCs) or Data Collection and Product Centres (DCPCs) within WIS. Secondly, developers of CDMSs will, where appropriate, build facilities into the CDMS which will allow the system to act as a provider of climate data via the WIS protocols. Some development work along these lines has already been carried out.

Data access and product development

Each NMHS should have a data access policy which would *inter alia* describe what data are available to users free-of-charge. Free access to data from a national station network differs from country to country. Basic metadata on the station network, giving information about location, geographic coordinates, altitude, sensor equipment and data availability, etc, should be provided, preferably, in a table and map view and without any charge from the NMHS. Even in Europe, where the NMHS policy is marked by very limited free access to data and station information, there is a move towards providing a greater amount of information without restriction. The project UNIDART (uniform data request interface) led by the German NMHS (Deutscher Wetterdienst), is an example of this.

There are many possibilities for providing user access to a database, including through SQL (Structured Query Language) if you use a relational database management systems (RDBMS), or other tools that help the user to communicate with a database (spreadsheet, shareware or

commercial applications). The most common way is to use applications that are able to handle an Internet browser. Then you can use the same applications for Intranet and Internet, while permitting appropriate access rights.

Staff who are responsible for providing data should understand how data have been treated – from source to the archiving stage. In earlier times, data management staff were responsible for developing products and the needs of external customers were often not sought. More recently, data provision has become responsive to the demands of end users, as well as the marketplace, and data provision services have developed in two broad directions.

Firstly, there are value added products that produce ‘customer-tailored solutions’. With these products the NMHSs earn revenue and the customers receive information tailored to their precise needs. These customers are generally not interested in additional but irrelevant information.

Secondly, there are bulk products, which are provided with less effort and a less selective coverage. Such products tend to be publicly funded and disseminated by less costly means. For example, the NMHS may publish them on its web site. In all cases, products should be presented in a professional form, because the NMHS will be judged by the quality of its products.

Data administration and monitoring

As has been emphasized in Section 3.1, data managers need to ensure that their systems and data are managed in a way that satisfies user (or business) needs. Consequently, systems and processes need to be put in place to check that the data being managed conform to the expectations of users. A useful way of doing this is to establish and monitor a carefully selected set of performance indicators that reflect the standards expected by data managers.

Data Managers must assume that there will be times when data do not flow from instrument through to the database and so they need to administer the data flow processes and closely monitor what data are collected in the database. Data flow problems should be identified and corrected before there are any consequences for climate data users. Responsibility for this sits firmly with data managers although other parties may be responsible for corrective actions (e.g. missing observations may be due to problems with communications technologies and so an IT section may need to take corrective action).

Section 3.4 describes important attributes of performance measurements in the context of more general database management. The general philosophy is the same when the focus is the climate data, i.e. any monitoring indicators developed should be associated with target values which allow for measurement of performance against user needs. Again, the indicators should be selected to be Specific, Measurable, Achievable, Realistic and Time-bound (i.e. SMART).

So what needs to be monitored? Looking at this from a user perspective might encourage a data manager to develop processes and monitoring indices to ensure that data in the database are:

- Appropriate in terms of types of climate parameters;
- Comply with the relevant national and international guidelines in respect of spatial density, frequency and length of record;
- Subject to appropriate quality control;

- Of acceptable quality;
- Made readily available in digital form; and
- Stored optimally in terms of security and accessibility.

Some of the above (e.g. climate parameters are appropriate) may require surveys to be undertaken or could be based on regular feedback from data users. The effectiveness of quality control procedures can be measured through, for example, false alarm rates, while the extent to which quality standards are achieved may look at the percentage of missing and erroneous data values. Indicators which track the proportions of data made available in digital and hard copy form may also be useful as measures of efficiency of data processing systems. Section 3.4 provides a more specific example of this. Monitoring digitizing rates also assists data managers in identifying efficiency problems and in planning workloads.

While care must be taken to ensure that these monitoring activities do not place a burden on the data management section, they play an important role in providing warnings of departures from acceptable performance. They can also be a useful motivating tool for staff.

Perhaps most importantly, the systems and processes in place should provide real-time alerts to business critical activities. This may include alerts to missing data during times where high profile climate products are produced (e.g. for government ministers) or where data are being provided to clients who are paying an urgency-fee for the service.

Finally, it is important that effective and efficient communication systems are in place to provide information to those who can take necessary corrective action.

Change management issues

Climate data are subjected to many and varied non-climatic influences and so the need for data managers to adopt good change management practices is necessary. Increased concern about climate change has elevated the importance of change management over the past couple of decades.

A good prerequisite to sound change management is to maintain good knowledge, metadata and other documentation on existing data management practices (see Sections 3.5 and 4.1).

The kinds of changes that need to be managed include:

- Changes to observation networks and systems;
- Changes to observational practices;
- The introduction of new data types; and
- Changes in the algorithms that compute derived data.

Transition issues between database management systems are discussed in Section 5.

Changes in observation networks and systems (e.g. site moves, replacement of sensor type) should ideally be reflected in the metadata records of affected stations and, if the changes are applied on a station-wide basis, in other documentation which is widely circulated and is readily accessible for future generations. Where a parallel observation program is in place, e.g. for a change in site, it will be necessary to manage the associated data streams (e.g. introduce a new site or maintain the old data in a separate database table). Changes to observational practices

should be maintained by the observations section of the NMHS and, again, the documentation should be widely circulated and readily accessible for future generations. The same applies to changes in the way that derived data are calculated although, while the observations section will be responsible for documenting changes to observation system-related algorithms, data managers will have responsibility for recording downstream changes associated with the database. Data managers should, however, attempt to present a consolidated list of the changes to users.

The introduction of new data types will usually mean integrating processes and systems required for storing, quality control, reporting, etc. For significant additions, it may mean taking on more staff or introducing new hardware or software and these often require considerable planning.

Consistent with quality management principles, NMHSs should consider the use of formal processes to help ensure the orderly management of change. WMO (2003a) describe efforts within the Meteorological Services of Canada (MSC) to do this. While the formal structures adopted by MSC may be beyond the resources of many NMHSs, the general principles are worth keeping in mind and in particular that:

- A group is established within the organization to oversee changes to systems, instruments, algorithms, processes, procedures and related documentation that influence the acquisition, processing, reporting and archiving of observations from the observational networks;
- Group membership covers the decision-makers associated with the aforementioned areas and scientific/technical support where necessary;
- Decisions are based on consensus to the extent necessary;
- A clear and transparent change request process exists;
- Change requests are reviewed and processed in a timely manner;
- The committee has powers to establish task groups to evaluate requests and make recommendations; and
- All requests, decisions and supporting documentation are made readily accessible.

Transition to a database management system

When tools currently in use for climate data management and data processing do not fully satisfy the needs of a NMHS, the question of acquiring a new database management system (DBMS) arises. The reasons could be diverse: to be able to handle larger data volumes than before, to answer new user's demands (e.g. delivering data through Internet), or because the current tools are not scalable enough or do not correspond to the present technology.

The choice of a system and the transition associated with a new DBMS is crucial because it commits a NMHS for a long period of time, i.e. the whole period during which the NMHS will use the chosen system. Putting forward a 'project approach', this Section describes the main phases (current situation analysis, functional solution design, and technical solution design) to select an appropriate system. Some considerations of database architecture (database design, data models) are also given for greater understanding and decision making. Finally the transition steps from the CLICOM CDMS to a new CDMS are detailed.

It should be stated at the outset that the fundamental message for making the transition between DBMSs is 'buy if you can, build only if you must'. To support this principle, the WCDMP has promoted the development of candidate CDMSs for countries to consider (See Appendix 1 and the case study presented in Appendix 2). This issue will be explored further below.

Choosing a climate database management system: what needs to be considered?

This Section uses the terminology of commercial business. Many NMHSs are funded wholly by government and therefore do not charge directly for their services; however, these NMHSs still need to demonstrate value for money to the funding departments by showing that services are delivered effectively to users. Hence the funding department is a proxy customer for the services delivered to the users.

Consider the business needs

The first stage when choosing a CDMS is to evaluate the business needs of the organization and the organizational environment within which the system will operate. These needs may be thought of as either internal or external. The external business needs will cover the existing customer and user requirements and expectations of future requirements. An analysis of the current and desired future products is needed. For example, if one of the major products is related to the yield of a particular agricultural crop, the CDMS might need to store data on date of emergence for that crop, and so the ability to specify such data types will be required.

What sort of changes to the business needs are possible during the life of the system? For example, an NMHS may need to develop more commercial revenue – this will require the development of new bespoke services, so ease of development of new products will be important.

The scope or boundaries of the organization need to be considered. This scope may be limited by the statutory framework for the organization. It will also depend on other co-operating or competing organizations.

For example, if the organization is an NMHS with responsibility for hydrology, seismology and the marine environment, then the range of data types will be wider than for an NMHS which only covers traditional meteorology. If there are separate organizations covering meteorology and hydrology, there will need to be agreement between them on responsibility for rainfall data, and also on the processes involved in collection, quality control and exchange.

In the case of an NMHS, there may be a requirement to store data from outside the national boundaries. Appropriate mechanisms will be needed to control use of this within WMO Resolution 40 (see Section 4.5) and any other agreements.

Designing the required system

To prepare the design of the required system, it is helpful to develop a problems/requirements list, to record the issues and to assign them a priority. This is best done as a collaborative process including all the stakeholders in the system. Of course, not all of the problems and requirements can be satisfied.

Money will usually be one of the stronger constraints, but staff availability and skills levels are also important. Hence the training needs of staff should be planned. Local availability of training and support for the hardware and software is very important. Use of hardware or software without local support should only be considered if timely and appropriate alternative arrangements can be implemented and sustained throughout the life of the system, and a greater level of redundancy will be needed.

The flow of data through the organization needs to be evaluated and changes planned. The four primary issues for a CDMS are control of the metadata, data entry processes, quality control processes and enquiry handling. The data entry may take place by a number of means, including direct ingestion from automated systems and key entry from paper forms (see Section 4.2.2). The details of how this is currently performed, and how this can be improved in the new system should be considered. The volume of data is important, both in terms of the routine ingestion and the total database size. Climatological data is typically held forever, so the backup and archiving strategy needs to be considered for a database which will grow throughout its life.

A particular issue for designing a CDMS is whether the system should also function as part of the synoptic operational systems.

Scalability

As part of the planning, scalability needs to be considered. If the organization changes, can the database change to suit? Typical issues will be the need to add extra data types, or to cope with large volumes of extra data: for example, a requirement to hold 10-minute data from AWSs. Generally the cost of hard disk storage is falling faster than the requirement for additional observational data is growing.

Is it possible to devolve the database, by producing reduced versions applicable to a district office containing a reduced set of the data applicable only to that district?

In general, keying data into the system as close to the point where it is first gathered has significant advantages in terms of the quality of the data. Devolving the database (or access to it) to local enquiry centres can have benefits in dealing with customers and users at those centres, through greater immediacy and improved understanding of customer requirements, as well as making better use of staff at those centres. The disadvantage is some loss of control and visibility, and greater support costs. It may be that a common architecture and CDMS would have benefit for other organizations, such as river authorities.

The potential barriers and costs of such changes in requirements should be considered as part of the long-term planning for the CDMS.

Architecture and technology

Once all of the problems and requirements have been established, the required system can be established to meet as many of them as possible.

For the hardware, the issues of reliability and maintenance are vital. In every country, PC technology is now well established, and hardware maintenance is available, at least for standard technology. Fault tolerance is very important so the mains power supply should be fed through a

regulating Uninterruptible Power Supply. For the server holding the database, the hard disks should be set up as a RAID (Redundant Array of Inexpensive Disks) array, preferably with at least three physical drives, so that the failure of one does not affect the data. If a system has to be available at any time or with a minimum of time stoppage then a redundant server could also be added. This redundant server will satisfy continuity if the main server fails.

Use of hardware architectures other than PC should only be considered where effective support is available to allow restoration after a failure within the time-scale allowed by the business needs (for a CDMS this will often be next working day).

Any CDMS needs to be protected by a system of backups. Ideally this should use a medium which can be read on other systems within the organization, typically writable CDs or DVDs, possibly also tape. The backups should be checked for readability on another machine at frequent intervals, and a cycle of backups including off-site storage should be set up.

For the operating system, similar considerations of cost, training and support apply. Microsoft Windows is well understood and supported commercially in almost every country, but is likely to suffer from viruses and similar malicious programs, and is open to unauthorized use. Hence there is also a need to invest in anti-virus software, and a firewall for external connections. Linux and other systems will typically require a greater level of expertise among the support staff; but if those skills already exist in the organization (as a result of other systems) then there can be useful savings.

The RDBMS can be a commercial product or an Open Source product. See section 5.3.3.4 for a detailed discussion of particular products. Availability and cost of training and support will be critical. The whole-life cost for the RDBMS will need to be estimated, and this can be complex. Particular issues are whether additional license fees will need to be paid when the hardware is upgraded, or if a devolved system (see Section 5.1.3) is set up in regional centres.

Client-server solutions may have significant advantages. Interfaces, like JDBC (Java DataBase Connectivity) or ODBC (Open DataBase Connectivity), are designed to communicate between databases and between users and database. This means that the main database server platform and RDBMS may be chosen for most efficient and safe operation, while multiple clients can be set up with the facilities specific to the business functions. Different client designs can be used for the data ingestion and QC tasks to those used for end-user queries. In particular, the client systems used for end-user queries must have considerable flexibility to quickly design and build new reports. Spreadsheet clients are useful where the extracted data is to be passed to end users as a data file. There are many tools (for example, Microsoft Access used as a client) which make the preparation of well-presented reports relatively quick and easy.

The quality of presentation of reports to end users, and the ease with which new products can be developed are very important to an NMHS. Good presentation tells the customer or user that the NMHS is a serious, effective and efficient scientific organization, and therefore helps to deliver the vision of the NMHS.

Choice of CDMS

Most of the available CDMSs, and the RDBMSs on which they run, are scaleable starting from a PC platform upward. As detailed above, the CDMS can be thought of as a combination of the

data model and a number of tools for doing the tasks required by the business needs. Ideally, one would test several CDMSs and then select the most appropriate one. An alternative is to develop a specific CDMS. In practice, the investment in time and effort is huge.

In most cases it will be effective to choose one of the existing CDMSs on the basis of the best fit to the business needs. It is useful to look at the experiences of similar organizations, but this is not always the best guide. The decision should include costings over the life-cycle, support, training and potential future developments. Then the CDMS should be adapted to local needs. Typical adaptations would be the addition of key entry forms for local use, automatic ingestion from other systems, and end-user reports. Such additional facilities may be useful to others running the same CDMS, and other groups may have good ideas that you can use.

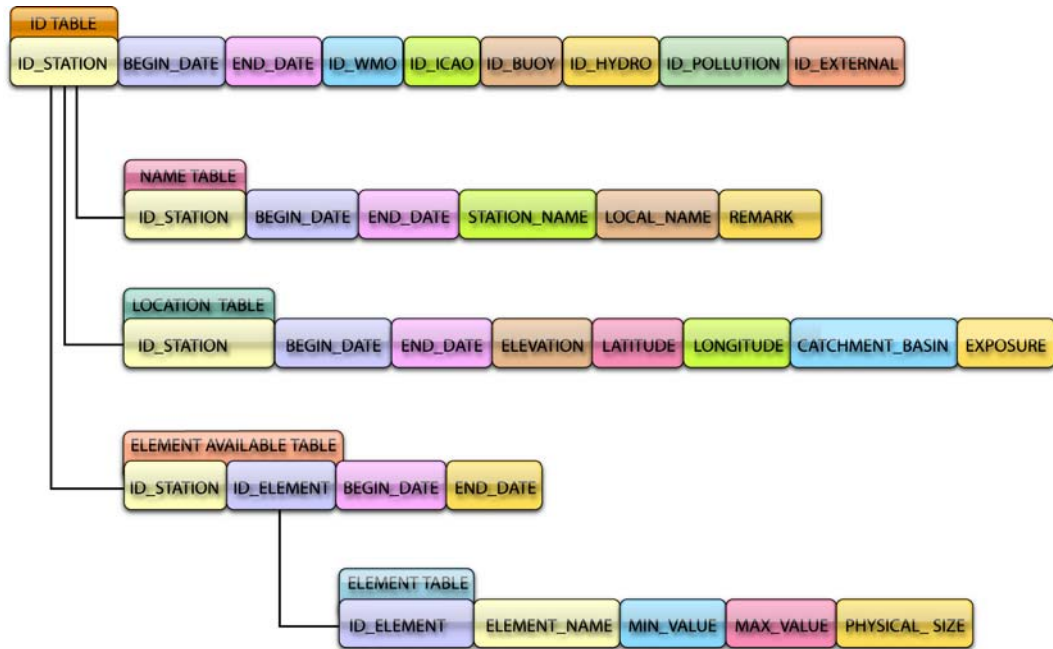
Database architecture considerations

This section brings the Relational Data Model into focus and leaves aside new model technologies (e.g. "object model" or "context model") that are, up to now, not commonly used by NMHSs.

A Relational Database Management System (RDBMS) is a collection of data represented in tables that are composed of rows and columns, in which all data relationships are represented by common values in related tables. For example, when storing information about the station's Identification, then a table "Id" could be defined, and one solution would be to decompose the Id information into several pieces of information, e.g. the WMO Id, the HYDRO Id, and the LOCAL Id as shown here:



The Id table could then be related to some other tables of the database, as shown here:



Defining a data model is the first stage for a database designer and consists of the optimization of the table structures: finding the best structure for each table and how the tables are interrelated in order to lay the foundations for an efficient system.

Database design considerations: normalization

To find those best table structures, a database designer has to follow certain rules, known as normalization.

Database normalization is a series of steps followed to obtain a database design that allows for consistent storage of data together with efficient access. Normalization is useful to:

- Limit data redundancy;
- Limit data loss; and
- Limit inconsistencies between data.

The exact design will be a compromise between performance and flexibility. Generally there is not a single best solution. This is the reason why the different CDMSs available have come up with different solutions.

Data models in use by CDMSs

Most CDMSs offer a Metadata model to store information on metadata and a data model to store climatological element values. The design approach for the Metadata model varies from a few related tables for some CDMSs (e.g. CLICOM) to a complex set of related tables for others. It all depends on the desired level of detail.

However, regarding the data model, some similarities are found. There are generally three data models used by CDMSs to represent climatological element values, each of them having advantages and disadvantages.

The Element Model

An Element Model represents data in tables, having, in each row, **different values of one element** observed at one station **at different times**.

For example, daily data could be stored in a Daily table. Each row would correspond to a specific station, a specific month, and a specific element. The attributes, i.e. each cell of a specific row, store the different values of that element and of that station for a given month (e.g. 31 values for January).

Table 5.1. Daily Data Table for the Element Model

Station's ID	Month/Year	Element	Value day 1	Value day 2	Value day 3	Value day 4	Value day 5	Value day 31
95123	01/2002	Tmin	23.4	25.2	28.3	26.5	27.8	24.9
66202	01/2003	Tmax						
								

Advantages: It is easy to add new elements; the data model remains the same even if a new element is added.

Disadvantages: Performance for real-time applications may be poor; many operations on the database can be more complex than would otherwise be the case.

The Element Model is used, for example, by CLICOM.

The Observation Model

An Observation Model represents data in tables having, in each row, the values of **different elements** observed at one station **at a given time**.

For example daily data could be stored in a Daily table. Each row would correspond to a specific station at a specific time. Each column of a specific row would store the values of the different elements observed.

Table 5.2. Daily Data Table for the Observation Model

Station's ID	Day/Month/Year	Tmin	Tmax	Rain	Min Humidity	Min MSL Pressure	Max Wind Speed	Max Wind Direction
33220	01/01/2002	24.5	33.4	0	72	1015.6	2.2	160
42500	01/01/2003	15.2	22.3	10.2	80	1013.4	3.3	210

Advantages: High performance for real-time applications; optimisation of data storage.

Disadvantages: Need to update the table structure if a new element that has not been included during the database design stage has to be added.

The Observation Model is commonly used by NMHSs - see the WMO Report from Benichou and Lee (1996).

The Value Model

A Value Model will represent the data values in tables having, in each row, **only one value** of **one element** observed at one station **at a specific time**.

For example, daily data could be stored in a Daily table. Each row would correspond to a specific station for a specific element and at a specific time.

Table 5.3. Daily Data Table for the Value Model

Station's ID	Time	Element	Value
33220	01/01/2002	Tmin	23.4
42500	01/01/2003	Tmax	16.3
22222	01/01/2003		

Advantages: It is easy to add new elements, the model is adaptable to a large range of data types.

Disadvantages: Optimization of data storage will not be done well, so this approach is not suitable for tables with huge amounts of data; also shares the disadvantages of the Element model.

In conclusion, each data model has its strengths and weaknesses and each one could be useful to a certain type of climate application. Furthermore, it should be possible to include, within the same database, data tables having different data models in order to respond to specific needs and constraints. For example, a CDMS could include an observation model for the hourly and the daily data and a value model for the daily pollution data.

Computer hardware and software considerations

Before defining the components of a new system in detail, a current situation analysis is recommended that covers staff expertise and hardware/software available, as well as a desired requirements study.

Current situation analysis

In order to decide which hardware or software is suitable for a NMHS, it is necessary to have a full knowledge/understanding of the current NMHS data management environment. Technical information coupled with expertise in data management is essential for this step and will help to get to a functional solution design. The key steps are:

Complete inventory and description of the hardware and software components currently available: computers, network, operating systems, DBMSs, applications in use, etc.

Current telecommunication possibilities in the country and/or the region: International and national telecommunication lines available: Internet, GTS, telephone, radio, etc.

Current expertise level of each NMHS agent on hardware/software: on hardware operation, on maintenance, on repair, on Operating Systems, on DBMSs, on programming languages, etc.

Current in-country expertise, or nearest location: of suppliers or other related organisations, their competencies and intervention capacities (e.g. maintenance, repair, spare parts).

Functional solution design: size up the requirements

Once the current technical capabilities and available expertise are described, a second step consists of having a precise description of the NMHS's requirements. Usually, the planned system should, at least, meet the useful functionalities offered by the 'old' system in place and act in accordance with the NMHS climatological practices. But furthermore, it is a good opportunity during a process of acquiring a new CDMS to take time to think about a possible re-definition of requirements. For example, the following are particularly subject to enhancement: data flow from the station to the database and then to the NMHS's users, delay of data acquisition and product delivery, data security, data access, data production. Those requirements will have a direct influence on technical choices (e.g. network, hardware/software, CDMS) for the new system.

The following questions are suggested in order to assist the person responsible for the description of the NMHS's requirements.

What usage and which functions?

Should the system store a single climate database? Or should it store some other databases (e.g. seismic or administrative records for the organisation)?

Are specific telecommunication lines needed? GTS, Internet, telephone, AMSS, etc.

What are all the functions that the system has to handle, especially with regard to the following domains: data acquisition, data access, data management and data processing, data production and data security.

Please refer to <http://www.wmo.ch/web/wcp/wcdmp/cdmsfuture/html/evaluation.html> for a complete list of criteria.

Which constraints?

The new system will have to be implemented in a specific environment, i.e. within the organisational context of the NMHS and its constraints, and therefore the following aspects have to be taken into account.

Which users?

How many users need to access the CDMS? Where are they located? Do they need real-time access to the data or is remote access sufficient? Have the users specific roles? The answer to this last question could help to determine if a distributed (e.g. both national and regional databases) or a centralised database (a unique database) is preferable.

What volume of information?

What is the current data volume requirement? What is the estimate of the required volume for the next 5 years? The answer to that question should be linked to the NMHS plans for the future: number of stations, elements to be stored, period, time schedule acquisition (e.g. every minute, every hour, etc.), key-entry and data acquisition policy, and network expansion.

Functional interactions: Internal, national, and international?

Drawing up functional schemas between the different NMHS's stakeholders within a data management operation including internal, national and international levels are especially useful. It facilitates a clear representation of all of the interactions and assists in the correct design of a network. Data flow from the observation location to the products intended for the users is particularly suited to this kind of schema. Such design could be extended to other topics like functional schema on the desired data operation or on the desired data access.

FUNCTIONAL SCHEMA

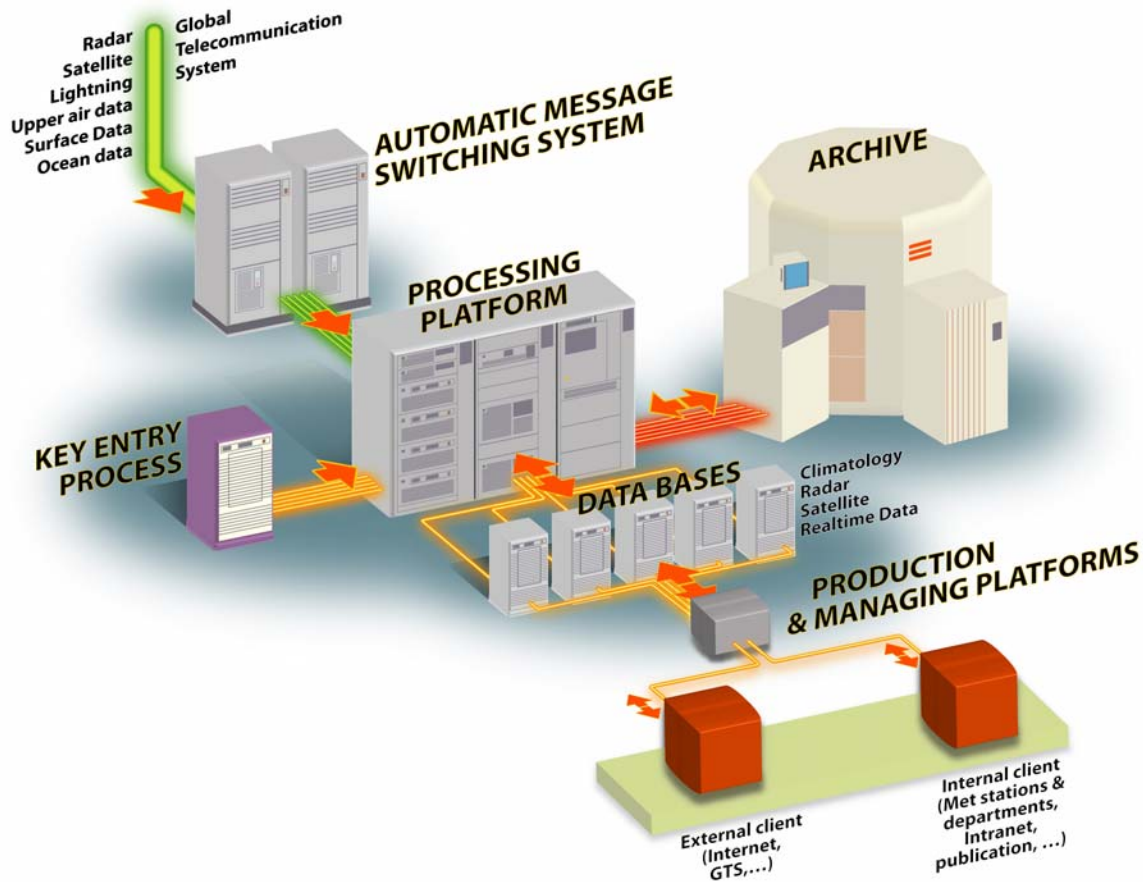


Figure 4. Example functional schema showing data flows and the various interactions at play within a data management operation.

Availability?

Are system stoppages allowed? Is there a minimum time requested by the system in order to switch off properly without any data loss or damage?

Compatibility?

How compatible will the new system be with existing hardware and software? For example, do some existing programmes have to continue to be operational with the new system, or do certain existing current equipment have to continue to communicate with the new system.

The technical solution design

With the information above and relevant computing expertise, a NMHS should be able to design a technical solution. Some important factors to consider now follow.

Budget choices

The total budget of a system is not simply the cost of its different components. Some other services should be included, such as:

- Warranty;
- Training cost (e.g. operation, maintenance, repair);
- Running cost (e.g. consumable, communication lines);
- A maintenance contract; and
- A budget for maintaining and developing the system: hardware and software enhancement or upgrade, corrective and adaptive maintenance, punctual assistance, specific developments.

Computer types

It used to be common practice to distinguish computers by their size, from the modest ones to the high-performance ones: microcomputer (or personal computer, PC), minicomputer and mainframe. Sometimes they could be distinguished by their application domains: simple terminal, workstation or server. Today, the minicomputer, as well as the simple terminal, is difficult to find on traditional markets. They have gradually been replaced, by the microcomputer (Figure 5), which has had a high growth in speed, memory and capacity. Therefore it is preferable to concentrate on the microcomputer description and to assume that the mainframe is no longer adequate, because of its large size and excessive costs, for most NMHS requirements.



Figure 5. The microcomputer (or personal computer, PC) is the primary tool for data management operations in many NMHSs today.

Microcomputers could be from different types:

- **Desktop:** that is office-oriented;
- **Workstation:** more powerful than a desktop and designed for specific applications;
- **Server:** designed for business-critical applications and able to deliver information to a set of clients; and
- **Cluster:** that consists of multiple interconnected servers.

Please refer to the glossary to get a more precise definition of these terms.

So, which type of microcomputer should be selected?

As a database is a critical bank of information, i.e. part of the climate heritage of a country, it is essential to have a robust and secure system as a priority.

A server is more robust, more scalable than a desktop, but it is quite feasible to convert certain desktops into a CDMS Server if hardware cost is a constraint for a NMHS.

In Appendix 3, three solutions are given, each including a different type of microcomputer. Their comparative advantages and disadvantages are described.

Operating System: Which OS?

If the existing market for CDMSs is considered, and the microcomputer option is selected, then two kinds of Operating Systems are dominating: Windows and Linux.

The Microsoft Windows environment almost has the monopoly for the world user desktop. This operating system is a commercial package and provides a server environment for databases and applications. To discover the windows versions please refer to the site: <http://www.microsoft.com/windows/>.

The Linux environment, which is closest to the Unix OS, is generally distributed as an Open Source Software. Linux is commonly used for Web and database servers. The latest distribution versions (see <http://distrowatch.com/> for the major distributions) have progressed in recent years and now give friendly interfaces for desktop users.

The choice between these alternative will differ between each NMHS and will require answers to the following:

- What are the competences of the NMHS staff?
- Are the desired CDMS or the NMHS technical environments compatible with Windows and/or with Linux?
- What will be the real cost?

A third solution, implemented by many NMHSs, consists of having a mix of computers running some components over Windows and some over Linux, depending on the application. But this solution requires a large range of expertise in the NMHS.

Finally, for NMHSs, which are currently running other Operating Systems like UNIX, OS/400, MacOS, or some other OS dedicated to specific Mainframes or Minicomputers, the decision between continuing to use them or to migrate to Windows/Linux needs precise analysis.

DBMS: Which DBMS?

Most of the Database Management Systems (DBMSs), currently used by NMHSs to manage climatological data, are of relational type and are called RDBMS. Meanwhile, as technology is rapidly evolving, new types of DBMS appear like Object Oriented DBMS, XML DBMS or Multidimensional DBMS.

This document will focus only on RDBMSs and Table 5.4 summarises the most popular ones.

Table 5.4. Information on existing RDBMSs.

RDBMS	Company	Operating System	Open Source	Comment
Access	Microsoft	Windows	No	Easy to use and widely used by the general public. Limited in terms of volume. Not a real client/server DBMS but rather a file/server.
DB2	IBM	Linux, UNIX, Windows, OS400	No	Compatibility with OS400, a standard for numerous big companies.
FileMaker	FileMaker	Windows, MacOS	No	Easy to use but limited in terms of scalability.
Firebird	Firebird	Linux, UNIX Windows	Yes	Based on Interbase.
Ingres	Computer Associates International	Linux, UNIX, Windows and OpenVMS	Yes	Well known RDBMS, has been released as Open Source.
Interbase	Borland	Linux, UNIX Windows	No	Good choice to manage small and medium databases.
MaxDB (SAP)	MySQL	Linux, Unix, Windows	Yes/No	Dual Open source and commercial licence. Able to manage huge amounts of data.
MySQL	MySQL	Linux, Mac OS X, Unix, Windows	Yes	Very popular on Internet. Does not yet have all functionality of some other DBMSs.
Oracle	Oracle	Linux, Mac OS X, UNIX, Windows	No	Like DB2, a standard for companies having numerous constraints and huge data volumes to process.
PostgreSQL	PostgreSQL	Linux, Mac OS X, Unix, Windows	Yes	Robust DBMS, one of the main standards in Open Source.
SQL Anywhere Studio, Adaptive Server IQ	Sybase	Windows, Novell NetWare	No	A free version for Linux is distributed: Adaptive Server Enterprise Express.
SQL Server	Microsoft	Windows	No	Good choice for medium-size organization.

The following issues have to be considered in choosing a RDBMS: RDBMS functions, adaptability to the NMHS context, cost, compatibility (hardware, software), and expertise available in the NMHS.

System and service transition issues

During a transition to a new climate database, new tools may appear and also new tasks, which could have implications on the day-to-day work of the staff. The transition to a new system provides a good opportunity to review the security policy and work sharing within the organization.

System and software security

Security is one of the major issues that a database administrator has to face. As a worst-case scenario, a NMHS may lose its entire database if there are poor security provisions in place.

Some basic recommendations are:

Prevent any loss of data through backup: using Network Attached Storage (NAS) technology, tape technologies (DAT, DLT, LTO), CD-ROM or DVD media.

Prevent any loss of climate data with archiving: Using the same technologies as used for backup may be possible. Keep three copies of the same archive in different secure places – and in different towns/cities if possible. Store the climate data in an easy readable format (e.g. ASCII).

Protect your server room: with access security, fire-security and lightning-security. Respect the safety notes in the operating manual for all the hardware devices (humidity, temperature, dust, etc).

Prevent any disk failure: with disks clone or appropriate technology (RAID).

Power supply: Add an uninterruptible power supply (UPS) to the system, giving enough time to stop the system appropriately. Investigate the possibility of solar energy in the case of frequent power supply failure. Check the earth electrode of the building and the electrical lines with adequate fuse wire.

Train the staff on security: and ensure relevant documentation is in place.

Computers: add redundant power supplies and redundant cooling on critical computers. A redundant computer system should also be considered.

Software and network security: Add firewall, anti-virus and anti-spyware software and ensure that these are regularly updated. Give each user a profile with specific rights.

Material: plan consumables for a sufficient period and check the availability of spare parts.

Day to day operation

Process and responsibility: Each process should be described and should be under the supervision of an identified person. If necessary, a process could be divided into sub-processes with its responsible agents identified. It is therefore quite common to find, in an IT division, a System Administrator responsible for operation and maintenance of the operating systems and for hardware and also a Database Administrator who carries out the functions of designing, implementing and managing the database operations. A Network Administrator would focus on making sure the NMHS 's LAN is tuned for optimum performance and security and interacts properly with Internet.

Documentation: Each process, and eventually each hardware or software set-up and its historical modifications should be documented (see Section 3.5). This is especially the case for the backup and archiving processes or any modification of the server's operating system that are critical for system survival.

Making the transition from CLICOM

This Section largely concerns NMHSs running a CLICOM system and that have decided to migrate to a new CDMS. However, parts of it will also be relevant to NMHSs who have an existing collection of climate data in some other format (e.g. on spreadsheets) and wish to migrate these. It is assumed that NMHSs have analysed their current environment (see Section 5.3.1), have come with a solution system design (see Section 5.3.2) and have already planned the desired hardware and software, including the DBMS and the CDMS (see Sections 5.1, 5.2 and 5.3).

What are the procedures required to make the transition?

Level of expertise required

To carry out the transition process from a CLICOM system to another CDMS, both CLICOM expertise and skills on the chosen CDMS are required. It is first important to fully understand the climatological rules in place on both systems to ensure the consistency of the transition. It is particularly the case for the quality flag definitions and the methods in use (data control, data generation) that have to reflect, at the end of the process, the NMHS practices. Second, expertise on DBMSs is also required. Among others, know-how on data models and their particularities (referential integrity, primary key, constraints, etc.) have to be mastered in order to properly import all CLICOM data into the selected CDMS. Some CDMSs may already have a CLICOM import facility, which makes this step easier.

Technically, transition methods are diverse and could use different tools like the CLICOM export functions, the DataEase Query Language so called DQL (which is the CLICOM DBMS), and also some DBMS import functions (as SQL loader) or standardized interfaces between databases.

Secure the existing CLICOM system

It is important to keep the current CLICOM system (hardware, software and climate data and metadata) in a safe place and operational, until the new system is proven to operate to the desired performance. During the transition time, both systems - CLICOM and the new system – may need to run in parallel for a period of time.

Prepare the metadata to be imported

Metadata are crucial for a complete understanding of climate data and have to be well managed. Generally, within a CDMS, metadata needs to be present before entering any climatological

values. Such information is stored in two places within CLICOM: the station history and the data dictionary.

One solution could be to export the CLICOM metadata in ASCII format and then to import this into the new CDMS. The import procedure will depend on the chosen CDMS and on its implemented metadata model. Some specific data processing will probably be required at this stage, depending on the degree to which the CLICOM data structures match the new CDMS.

Migrate the station history information

This part contains both current and historic information on station location, observing practices and instrumentation. It is stored in 3 different DataEase forms: STN GEOGRAPHY, STN OBSERVATION and STN ELEMENT. Pieces of information contained in those forms could be exported into ASCII files through using simple DQL reports.

Naming conventions on elements, or codes (e.g. in CLICOM an active station has an end-date equal to "9999-12-31") will have to be replaced with conventions used in the new CDMS.

Migrate the data dictionary information

In particular, this second migration involves the:

- Element definitions, in the ELEMENT DEFINITION DataEase form, that defines each climatological element managed by CLICOM;
- Archive CLICOM description, in the DATASET INFORMATION DataEase form; and
- Quality flag definitions used by CLICOM in the MISC CODE DEFINITION DataEase form.

Particular attention has to be given to the possible differences between the CLICOM quality flags and those of the new CDMS. Some of the main transition issues are whether:

- The results of all quality controls performed with CLICOM, i.e. data value plus its quality flag information, have been preserved; and
- A certain loss of information may be unavoidable and so the NMHSs must prepare to manage this risk.

Prepare the climate data to be imported

Conserve the quality flag information

The meteorological information associated with each value in the CLICOM database, i.e. the quality flag, has to be retained as far as possible. This is especially the case for dubious values (D quality flag), estimated values (E quality flag), generated values (G quality flag), results based on an incomplete period (I quality flag), missing values (M quality flag) and trace amount of precipitation (T quality flag). In losing this information, a NMHS would have to:

- Re-quality control all values within the new CDMS which would very likely involve a significant amount of work; and

- Refer back to original data, which could be the 'paper archive', in order to recover some information like trace amount of precipitation.

Importing the CLICOM data

CLICOM data are stored in the following DataEase forms:

- MONTHLY DATA form, for monthly element values;
- TEN DAY DATA form, for 10 day element values;
- DAILY DATA form, for daily element values;
- SYNOPTIC DATA form, for 3 hourly element values;
- HOURLY DATA form, for 1 hour element values;
- FIFTEEN MINUTE DATA form, for 15 minutes element values; and
- UPPER-AIR DATA form, for upper air element values.

During the data import to the new CDMS, the missing values, with an "M" for quality flag and "-99999" for value, should be considered with care as these have two meanings in CLICOM. These are:

- No values because no observation was made at that time (e.g. because of instrument malfunction); and
- No values because this day does not exist (e.g. in CLICOM, values of the 31 of February are flagged as missing).

Depending on the data model of the new CDMS, a NMHS could export its climatological data from CLICOM into two different ASCII formats, which are the:

- Traditional CLICOM export (with the CLICOM menu # 2.2.3) that will produce an ASCII file with an Element structure; and
- Observation CLICOM export (with the CLICOM menu # 2.2.4) that will produce an ASCII file with an Observation structure.

Collect information on CLICOM quality control set up

Methods of quality control, which have been performed on the CLICOM database, have to be recorded as part of the new CDMS metadata. They trace the history of the NMHS quality control and may prove to be useful for further enhancing the associated algorithms or simply to renew the same techniques with the new CDMS.

The CLICOM quality control set up is stored in files 'ELEMCHKS.*' (ELEMCHKS.DLY for daily data, ELEMCHKS.HLY for hourly data, etc.) under the directory CLICOM/DATA.

Intensive testing of the new system

Finally, thorough tests have to be made on the new system in comparison with the old system. Verifications should make sure that any loss of information is acceptable for the organisation and that the new system respects the official practices of the NMHS and also the client's expectations.

Sustaining data management operations

Resource requirements, including staffing

As discussed in Section 5 above, one of the first questions faced by a NMHS establishing a CDMS is 'buy or build'. In either case substantial resource commitment will be required, more particularly if the choice is 'build'. If the choice is 'buy', it is the choice of software which will determine, or at least influence, the selection of hardware. Hardware can then be selected in the light of such a constraint, bearing in mind also issues such as the volume of data to be managed, the available budget, and the degree of robustness required.

An important issue to be considered is that of version management. Whatever system is selected, it will be important to manage the versions of all the software components. Unfortunately, many problems can arise involving incompatibilities which eventuate when inappropriate versions of software are used in conjunction. For this reason, in the specific case where open-source software is used, it may still be desirable to subscribe to the services of an open-source provider, such as Red Hat™ or one of their competitors, to reduce the risk of problems in this area.

Establishing a system for climate data management has considerable staffing implications. The type of staff required fall under a number of headings, Firstly, in the order of data flow (and assuming that the observers are already in existence), we have:

- Data preparation operators who are involved in keying data which arrives on paper forms;
- Meteorologically-knowledgeable staff who can carry out QC work and interpret meteorological needs as necessary within the data centre (e.g. former observers or those who can assist in the identification of requirements and design of systems such as new QC systems); and
- (Probably) staff with a good knowledge of the database, to carry out the more complex data retrievals;

'Behind the scenes' staff are also needed (who don't necessarily need any detailed meteorological knowledge):

- Database administration staff;
- System administration staff; and
- If the NMHS wishes to make its own country-specific extensions to the system then IT staff will be needed to maintain and extend the application software. Of course, if the NMHS chooses to develop its own system, a larger investment of effort by IT staff should be anticipated.

Training

Training requirements associated with climate data management follow naturally from the range of associated staff responsibilities. Users of a system all need training in aspects of that system that are relevant to them. For example:

- Data entry and QC staff will need to understand the ways in which the system will respond to the entry of apparently suspect or erroneous data, and what recovery action then needs to be taken;
- While it is possible that a system can be treated as a 'black box' from the data retrieval point of view, with data being available only in a small number of predefined formats and with predefined selection options, it is more likely that there will be a requirement for staff to be able to develop their own tailored retrieval processes. Assuming that the CDMS is based on a RDBMS (as all six of the currently-defined WCDMP systems are) it will be possible to retrieve data using the SQL language, so staff will need to have an understanding of its capabilities and limitations;
- Expertise will also be needed in the 'infrastructure' areas of operating system administration and database administration. In a small establishment these skills could be combined in a single (well-trained) individual. However, each of these is a substantial area in its own right, and in a larger organization it would be expected that there would be one or more specialists in each area; and
- IT staff will need to acquire some familiarity with meteorological issues, and this needs to be provided by appropriate staff of the NMHS.

Occupational health and safety issues

There are not many occupational health and safety issues that are specific to climate data management systems (although dust associated with old records in a data rescue exercise and the dangers associated with lifting weighty paper records, may be exceptions). However there are a number of issues that apply to most computer systems and which should be kept in mind. These include:

- Ergonomic issues associated with work stations, keyboard positioning, furniture etc – inappropriate furniture and work practices can lead to a wide variety of physical problems;
- Possible radiation hazards from the back of monitors;
- Simple physical hazards which arise from having many cables and power leads lying around, as tends to happen in computer installations; and
- Stress, which is a common problem in many computer installations.

The organizations' occupational health and safety practices will be governed by legislation in many countries and these practices and any associated responsible officers should be consulted for further information.

Conclusions

These guidelines are intended to provide NMHSs with information on best practice climate data management with some emphasis on building knowledge and capacity in:

- The phases and stages of managing computerized data;
- Available database technologies and selecting an appropriate database;
- Making the transition to a modern climate database system; and
- Sustaining data management operations.

There has been a deliberate focus in these guidelines on those aspects of climate data management that are of interest to NMHSs wishing to make the transition to a modern climate database management system and, just as important, on what skills, systems and processes need to be in place to ensure that operations are sustained. This document is designed to assist NMHSs in evaluating the future strategic direction to be taken with regard to their needs in managing climate data.

From the outset, it is essential that both the development of climate databases and the implementation of data management practices take into account the needs of the existing, and to the extent that it is predictable, future data users. Data managers also need to be mindful of the principles for long-term climate monitoring. The ability to maintain the continuity, homogeneity and, ultimately, quality of climate data is greatly influenced by how well observation networks and systems are managed.

It is also essential to be able to store quality controlled climate data in a readily accessible format in order to support the data requirements of climate programs within the NMHSs as well as support various WMO programs. The climate record is a resource for the present and the future and it is important to safeguard this information for future generations.

One of the key technological decisions to be made is whether to take an existing CDMS design, and its associated software, or develop a tailored CDMS using one of the RDBMSs that are available in the world today. In all likelihood this is likely to be a decision that is made for financial reasons since the development of a CDMS and its associated software can require a large initial up-front investment. However, the decision making process also needs to take into account issues such as data flow and quality control as well as technological factors. It is strongly recommended that NMHSs consider the 'buy' option very carefully as the 'build' option requires far more resources and will usually be far more costly.

Acknowledgments

The authors wish to thank Amir Delju (WMO) and Suresh Boodhoo (President CCI) for their support of this work. Thank you also to Rod Hutchinson and Mike Macaskill (both Australian Bureau of Meteorology) who provided reviews of this document.

References

Benichou, F. and Lee, D.M. 1996. Use of Relational Management Systems in Climatology in WMO National Meteorological Services, WMO Report. World Meteorological Organization, Geneva.

Date, C.J. 1999. An Introduction to Database Systems (7th Ed). Addison-Wesley, 938pp.

GCOS. 2003. Second Report on the Adequacy of the Global Observing System for Climate. GCOS Report No. 82 WMO-TD 1143. World Meteorological Organization, Geneva.

<http://www.wmo.ch/web/gcos/gcoshome.html>

WMO. 1995. Manual on Codes: International codes, Volume I.1 (Part A). World Meteorological Organization, Geneva.

<http://www.wmo.ch/web/www/DPS/NewCodesTables/WMO306vol-I-1PartA.pdf>

WMO. 2002. Report of the CLICOM-DARE workshop; Report of the international data rescue meeting, WCDMP-No.49/WMO-TD No.1128. World Meteorological Organization, Geneva.

http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-49/WDMP49_menu.html

WMO. 2003a. Guidelines on Climate Observation Networks and Systems, WCDMP-No.52, WMO-TD No.1185. World Meteorological Organization, Geneva.

<http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-52.pdf>

WMO. 2003b. Guidelines on Climate Metadata and Homogenization, WCDMP-No.53/WMO-TD No.1186. World Meteorological Organization, Geneva.

<http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-53.pdf>

WMO. 2004. Guidelines on Climate Data Rescue, WCDMP-No.55/WMO-TD No.1210. World Meteorological Organization, Geneva.

<http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-55.pdf>

WMO. 2005. Guide to Climatological Practices (Third Edition). World Meteorological Organization, Geneva. (In preparation).

<http://www.wmo.ch/web/wcp/ccl/guide/guide.3e.shtml>

WMO. 2006. WMO Information System.

<http://www.wmo.ch/web/www/WISweb/home.html>

Glossary and Definitions

ACCESS	A relational database from Microsoft. Usually run on a local workstation although it can be installed on a server and shared.
Antispyware	A computer program designed to detect and protect from unwanted programs known as spyware.
ASCII	American Standard Code for Information Interchange. A de facto standard for the codes used to represent numbers, letters, punctuation marks and other symbols in a computer.
Byte	The smallest addressable unit in a computer. In an 8-bit computer it can have 2^8 or 256 values.
CDMS	Climate Database Management System. A database designed to manage climate data.
Central Processing Unit	The main controlling processor in a computer. Usually described in terms of the number of processing cycles per second and the band width of it's BUS. For a given configuration, the faster the CPU the faster will be the operations and response of the computer.
CLICOM	CLimate COMputing project. – A climate database developed under the auspices of the WMO and designed to run on a DOS or Microsoft Windows operating system.

Client/Server Database	The database is stored on a single server but applications using the data are executed on multiple client computers.
CLIMAT	An international computer exchange format for climate information from defined surface observation sites. These messages are generated and exchanged on a monthly basis between members of the WMO.
CLIMAT TEMP	Similar to CLIMAT messages but include data from upper air observations.
Climate data	Data used to describe or understand the climate system, which may include data corresponding to physical, chemical and biological properties, representing atmospheric, oceanic, hydrologic, cryospheric and terrestrial processes. The attributes of continuity and homogeneity are generally given high priority.
Climate data management	The Management of climate data in order to provide reliable quality data to users. Usually stored in a computerized database. The term can include the tools that are used to provide climate products to the customer.
Cluster	See 'Servers'. Clusters consist of multiple interconnected servers typically with a front-end server handling all the user interactions. Clusters are scalable, making it relatively easy to increase their capability.
Computer type	General classification of computers based upon size and performance. Types range from the most powerful, multi-user and large mainframe computers through to the middle of the range minicomputers and to the single user personal (or desktop) computer.
CPU	See Central Processing Unit .
Database Model	Refers to the way in which the database is designed. There are a number of models which can be followed of which the Relational model is the most common. They include Relational Model, Object/Relational Model, Object-Oriented Model, Semi structured Model, Associative Model, Entity-Attribute-Value (EAV) data model, and Context Model.
Data rescue	An ongoing process of preserving all data at risk of being lost due to deterioration of the medium, and the digitization of current and past data into computer compatible form for easy access.
Database	A collection of information organized in such a way that a computer program can quickly select desired pieces of data. The information within a database is usually interrelated.
DataBase Management System (DBMS)	The set of software used to develop, implement, manage and maintain data stored in a database.
Desktop	A general term for the backdrop on a computer screen where icons representing programs files and disks are displayed and from which applications can be launched.
Desktop computer	An independent general purpose computer capable of being used on its own. As the name implies it sits on a desk rather than being portable like a notebook (laptop) computer. Desktop computers are usually modular and do not have a single configuration.
Distributed Database	A Database where the Database Management System is distributed across multiple nodes or processors but managed as a single entity.

Double Keying	A method of data entry designed to minimize keystroke errors. The same data is keyed in by two different keyboard operators. Based upon the theory that the probability of both data entry persons making the same mistake is low.
Excel	A spreadsheet program developed and marketed by Microsoft.
Firewall	Software designed to prevent unauthorized access to or from a private network (This could be a single PC). In the case of a network it resides on a gateway server protecting the network from users in other networks.
GB	See Gigabyte
GCOS	Global Climate Observing System . GCOS is intended to be a long-term, user-driven operational system capable of providing the comprehensive observations required for monitoring the climate system, for detecting and attributing climate change, for assessing the impacts of climate variability including climate change, and for supporting research toward improved understanding, modeling and prediction of the climate system.
Gigabyte (GB)	Used to describe the storage capacity of computer devices. A gigabyte is $1024 (2^{10})$ Megabyte.
GSN	GCOS Surface Network – a baseline GCOS network for surface climate data.
GTS	Global Telecommunication System . A global telecommunications network operated by members of the WMO to exchange meteorological and related data and products.
GUAN	GCOS Upper Air Network – a baseline GCOS network for upper air data.
GUI	The Graphical User Interface of a computer screen. A system based on displaying various windows and from which various windows controls such as buttons, check boxes, menu bars and list boxes are used to initiate action.
Hard disk	A permanent storage device in a computer consisting of one or more platters with magnetic surfaces. Space on a hard disk is reusable. The operating system and other software are typically stored on the hard disk.
Hard drive	See Hard Disk .
MacOS	The operating system used by Apple in its Macintosh computers.
Mainframe	Generally acknowledged to refer to a large high powered computer capable of supporting multiple users (often hundreds) and running multiple programs simultaneously. Mainframes are usually housed in a secure and controlled environment. The difference between low-end mainframes and high-end minicomputers and servers is becoming smaller and less well defined.
MB	See Megabyte .
Megabyte (MB)	Used to describe the storage capacity of computer devices. Usually taken to mean 1,048,576 bytes (1024^2 or 2^{20}). Occasionally it can mean 1,000,000 bytes or 1,024,000 bytes depending upon its usage.
Metadata	Metadata is data about data. It is information about the data of interest. For Meteorological data this would include the type measuring instrument, information about the siting of the instrument, etc, as well as information about the content and format of datasets.
Meteorological	Another term for a meteorological datum or meteorological element

Parameter	of interest such as temperature.
Microcomputer	See Desktop computer and Personal Computer. Another name used for personal computers or desktop computers.
Minicomputer	A class of computers between mainframes and microcomputers. The term arose when there were distinct differences between mainframes, minicomputers and microcomputers. This is becoming less the case with time. Often used to refer to servers and high end workstations.
MOV	M ean O f V erification.
NMHS	N ational M eteorological and H ydrological S ervice.
Normalization	A process used in the design of tables in a relational database in order to develop an appropriate structure which minimizes unnecessary redundancy. Five normal forms have been identified.
ODBC	O pen D ata B ase C onnectivity - An open standard for accessing different database systems. For instance, accessing Oracle from Microsoft Excel or Access.
Open Source	A term to describe computer software whose source code is freely available for others to view and modify. This means that the software is available free of charge (or sometimes for the cost of distribution).
Operating System	The system software in a computer responsible for control and management of hardware, system operations and processes. The operating system controls the operation of application software such as a word processor. The most well known operating systems for personal computers are Microsoft Windows, Mac OS (Apple computers) and Linux.
OS/400	OS/400 is the operating system used by IBM for their AS/400 series of computers.
OVI	O bjectively V erifiable I ndicators.
Personal Computer or PC	An independent general purpose computer capable of being used on its own. This term is often used just for IBM or IBM clone computers running Microsoft operating systems but it can refer to any standalone general purpose computer.
Platform	In computer terminology this refers to the combination of hardware, software and operating system that is being used. Often used to refer just to the operating system.
Processor	Usually refers to the Central Processing Unit in a computer. See CPU.
QA	Q uality A ssurance.
QC	Q uality C ontrol.
RAID	R edundant A rray of I ndependent D isks. A grouping of data storage disks in a single unit which provides redundancy when any single disk fails and so reducing the risk of losing the data on that disk. Typically data on any one disk is mirrored to one or more of the other disks.
RAM	R andom A ccess M emory. Computer storage which provides faster access to the data being stored. It can be written to and read from but does not retain the data once the computer is turned off. In general the more RAM a computer has the better it will perform.
RBCN	R egional B asic C limatological N etwork
RBSN	R egional B asic S ynoptic N etwork
RDBMS	R elational D ata B ase M anagement S ystem. Often used

interchangeably with Relational Database but strictly refers to the system used to manage the database.

Relational Database	A database where data is stored in related tables. Data from different tables can be retrieved or manipulated because there is a defined relationship between the tables.
Server	A computer that can be accessed and used by more than one user over a network. Servers are typically scalable enabling a Server to be configured to meet the needs of an organization and allowing for it to grow to meet future requirements. A server can have multiple processors and peripheral devices such as hard drives.
SMART	Specific, Measurable, Achievable, Realistic and Time-bound, which are desirable properties for measuring database performance.
Spyware	A general term for unwanted programs that are usually installed without the knowledge of the user through the internet. These programs monitor actions on the computer and gather data about usage. These programs can include remote control programs and symptoms include pop up advertisements, a general slowing of the computer and unexpected toolbars on web browsers. See also Adware.
SQL	Structured Query Language – a computer language designed for querying and retrieving data from a database.
UPS	Uninterruptible Power Supply - A backup power unit that provides continuous power when the normal power supply is interrupted. Ensures that the computer remains operational allowing for a structured closure of the computer should the power outage be significant.
WMO Workstation	World Meteorological Organization. A high end general purpose computer that generally is more powerful and more robust than a desktop computer. Often used for engineering or graphical applications requiring large amounts of memory and a fast processor. Can have more than one processor and hard drive.

Appendix 1 CDMS descriptions

This is essentially from <http://www.wmo.ch/web/wcp/wcdmp/cdmsfuture/html/evaluation.html> . Short paragraphs or web links are provided to other known systems.

Numerous expert meetings, together with the responses to a questionnaire sent to all WMO members, six NMHSs proposed a CDMS as a potential successor to CLICOM. The **Czech Republic** with the *CLIDATA* system, **France** with *CliSys*, **Jordan** with *JCDMS*, **Russian Federation** with *CLIWARE*, **Tunisia** with *SDCLIM* and **Zimbabwe** with *ClimSoft* (Table A1.1).

Following the recommendations of the WMO Fifty-third Executive Council and the 13th session of the Commission for Climatology (November 2001), a workshop was held at the WMO Secretariat from 27 May to 1 June 2002 to evaluate those six CDMSs by an external team of experts. A report for each system along with a comparative criteria table was produced and is available at this address: <http://www.wmo.ch/web/wcp/wcdmp/cdmsfuture/html/evaluation.html> .



Figure A1.1. Evaluation team for the WMO Climate Database Management System (CDMS) project.

Table A1.1. WMO's Climate Database Management Systems (CDMSs)

CDMS	NMHS	Link	DataBase Management System
CLIDATA	Czech Republic	http://www.clidata.cz/	Oracle
CLIMSOFT	Zimbabwe	http://www.met-elearning.org/moodle/course/category.php?id=4	Access, SQL/Server
CLISYS	France	http://www.mfi.fr/productsandsolutions.htm	Oracle, PostgreSQL
CLIWARE	Russian Federation	http://cliware.meteo.ru/meteo/index_en.html	Oracle, MaxDB, etc.
JCDMS	Jordan	http://www.jmd.gov.jo/	Oracle
SDCLIM	Tunisia	http://www.meteo.tn	Oracle

Beyond the WMO evaluation process, other CDMSs have been developed by NMHSs or private companies (Table A1.2).

Table A1.2. Climate Database Management Systems developed outside of the WMO CDMS project

CDMS	NMHS or private company	Link
AMS	Food and Agriculture Organisation	http://www.foodsecinfoaction.org/News/tools_cw_02.htm
CLDB	Microstep-MIS	http://www.microstep-mis.com/products/meteorology/cldb/english.php
QUALIMET	Ernst Basler + Partner GmbH	http://www.ebp-de.de/en/geschaeftsbereiche/information/projekte/165/

Appendix 2 Case study of CDMS implementation

In 1988, the minicomputer used for managing the climatology data of Madagascar broke down and was never replaced. The centralised management and processing of climatological data came to a halt at that time. Fortunately, a national project was undertaken to collect the data from the minicomputer system that were initially stored on tape, and to copy these to CD-ROMs.

In the 1990s, with the coming of microcomputers, the IT and Database Department resumed the entry of climatological data. To do so, it used some basic spreadsheet software. This situation was insufficient for gathering and sharing data in an appropriate manner within and between the different NMHSs and to secure the data. Therefore, Madagascar NMS drew up a project within the framework of the WMO Voluntary Cooperation Programme (VCP) with the objective of implementing a national climatological database and to set up a modern, standard climate data collection, storage, processing and dissemination system for a resource management system.

The benefits expected were:

- To permit more effective participation at a global level in implementation of World Climate Programme activities, especially with regard to studies of regional climate change under the influence of natural and human factors; and
- To improve the real-time supply of climate information to various sectors of the economy, which will help to protect and develop natural, human and financial resources.

France decided to finance this project and Météo-France was in charge of its management in offering the assistance of its agents. This project has been conducted in two stages.

The first stage, which included a study of the present environment, arrived at the system definition. This definition comprises a server, a microcomputer type (that meets high security standards) supplemented by two client stations, a printer, a scanner, archival and backup units, the consumables required and training manuals. The system is associated with the Oracle DBMS and runs with a Linux core operating system.

The integration of the system, its specific setup and operational testing began in September 2003 and lasted three months. It was undertaken in the premises of Météo-France in Toulouse, under the supervision of system administrators, an Oracle administrator and a climatologist. The system was accepted by the Direction of Meteorology and Hydrology, Madagascar (DGM) in March 2004, after which the installation and training mission could begin. That took place in the month of April 2004, over a period of 15 days, with the participation of specialists from Météo-France. It involved installing and starting up the system in the NMHS, integrating the historical surface data relating to Madagascar, and building up and training the teams of system administrators, CDMS administrators and data entry operators.

Stage 2 of the project involves consolidating the skills acquired, continuing the training and developing new procedures.



Figure A2.1. System demonstration to Madagascar NMHS services

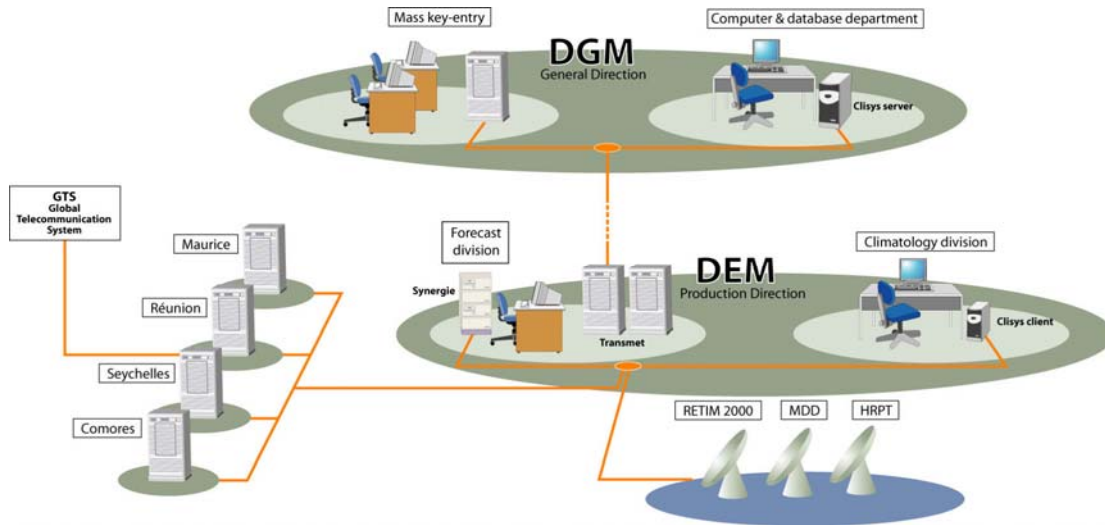


Figure A2.2. Madagascar system architecture. The CDMS system (CliSys) is connected to all the different services of the NMHS. Data from the GTS are collected via an AMSS (Transmet) that supply data for the forecast workstation tool (Synergie).

Appendix 3 Hardware examples

Hardware microcomputers solutions to manage a database

Three hardware solutions are given here in order to manage a climate database, from a simple desktop to a robust server. The system's cost, as well as the system's security and performance, increase from solution 1 to solution 3: desktop, workstation and server. Meanwhile, for those three solutions, particular attention has been given to security by mixing the different hardware security configurations and by adding, for each case, backup and archive possibilities.

Hardware security

Three methods for providing hardware security are presented.

The clone method consists of having the exact image of a hard disk, **at a given time**, on another hard disk - so called the clone. When the administrator assumes that the system is running properly, a clone of the hard disk system is built. Then, later on, if errors occur on the system (e.g. hard disk failure, files lost, bad configuration) the administrator is able to get back to the situation where the system was working properly through using the clone. This kind of security is quite valuable for data that do not change too much, e.g. like the operating system and the software applications. It is not, however, sufficient for a database that accumulates data every day. In that case an archive and backup strategy should be added (solution 1).

The disk mirroring is a method of storage in which data from one disk is duplicated to another disk **at the same time** so that both drives contain the same information. This technique, also known as RAID level 1, provides redundancy in case of disk failure and could be supplemented with a clone security (solution 2).

The striping method will safeguard against data corruption through copying data to a set of disks. If one disk failed, all of its data are automatically rebuilt on the other disks of the set. This method is also called RAID level 5 technique (solution 3).

Archive security

Archiving could be defined as a method of storing data, for **as long and as safe as possible**. It is specially designed for climatological data which, when it has been subject to quality control, is hardly modified. The support media should stand the test of time and that is particularly the case of CDs, DVDs and NAS technologies. For climate data, it is recommended to archive into a format readable by most software, e.g. the ASCII format. The life expectancy of CDs or DVDs is not unlimited. At the end of every session it is wise to copy the information from one media onto another.

Backup security

Backup can be defined as a method to store information **at a particular time** in order to retrieve, in case of loss, the total system information or some pieces of information at that particular time. It is useful when, e.g. a daily system backup is in place and all daily backups are kept on a weekly basis. Under these conditions, it is possible to retrieve a file at the end of a week that has been lost or damaged on a particular day during that week.

Finally, if it turns out that one server is not sufficient to handle the climatological database of a NMHS, a cluster solution has to be considered. That consists of multiple interconnected servers and storage devices forming what appears to users as a highly accessible system.

Table A3.2. Three hardware solutions to climate database management

Function	Solution 1: Desktop	Solution 2: Workstation	Solution 3 : Server	Remarks
Operating system	Linux, Windows	Linux, Windows	Linux, Windows	Check the compatibility of the OS with all devices and all software in use.
Processor	1 processor	1 to 2 processors	1 to 128 processors	Refer to the requirements of the different software in use, the number of clients and the NMS constraints.
Memory	Between 250 Mb to 4 Gb	Between 512 Mb to 8 Gb	Between 512 Mb to 1000 Gb	Refer to the requirements of the different software in use, the number of clients and the NMS constraints.
Hard disk	Limited number of hard disks. Example: 2 hard disks, disk 1 : disk for all of the system, disk 2 : clone of disk 1.	From 1 to 4 hard disks. Example: 3 hard disks; disk 1 : disk for all of the system, disk 2 : clone of disk 1, disk 3 : mirror of disk 1 (Raid 1).	From 1 to 192 hard disks. Example: 6 hard disks, disk 1 : operating system, disk 2 : clone of disk 1, disk 3, 4 and 5 : database on RAID 5, disk 6 : spare disk.	Analyse the volume requirement. Choose the appropriate hard disk type: IDE, SATA, SCSI, and the appropriate architecture: clone, mirroring, stripping. Plan a spare disk.
Archive	CD-Writer	DVD-Writer	NAS	Analyse the most appropriate technology.
Backup	CD-Writer	Tape technology, NAS	Tape technology, NAS	Analyse the most appropriate technology.
Power security	UPS	UPS	UPS	Analyse of the most appropriate technology according to the constraints.
Software security				Fire-wall, anti-virus, anti-spyware and all updated. Backup and archive software.
Warranty				Ask for a warranty over a reasonable time period.

REPORTS PUBLISHED IN THE

WORLD CLIMATE DATA PROGRAMME (WCDP)/

WORLD CLIMATE DATA AND MONITORING PROGRAMME (WCDMP) SERIES

- WCDP-1 WMO REGION III/IV TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT AND USER SERVICES, Barbados, 22-26 September 1986 and Panama, 29 September 3 October 1986 (available in English and Spanish) - (WMO-TD No. 227)
- WCDP-2 REPORT OF THE INTERNATIONAL PLANNING MEETING ON CLIMATE SYSTEM MONITORING, Washington DC, USA, 14-18 December 1987 - (WMO-TD No. 246)
- WCDP-3 GUIDELINES ON THE QUALITY CONTROL OF DATA FROM THE WORLD RADIOMETRIC NETWORK, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 258)
- WCDP-4 INPUT FORMAT GUIDELINES FOR WORLD RADIOMETRIC NETWORK DATA, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 253. p. 35)
- WCDP-5 INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS, 1989 edition (WMO-TD No. 293)
- WCDP-6 CLICOM PROJECT (Climate Data Management System), April 1989 (updated issue of WCP-I 1 9) - (WMO-TD No. 299)
- WCDP-7 STATISTICS ON REGIONAL NETWORKS OF CLIMATOLOGICAL STATIONS (based on the INFOCLIMA World Inventory). VOLUME II: WMO REGION I - AFRICA (WMO-TD No. 305)
- WCDP-8 INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS - HYDROLOGICAL DATA EXTRACT, April 1989 - (WMO-TD No. 343)
- WCDP-9 REPORT OF MEETING OF CLICOM EXPERTS, Paris, 11-15 September 1989 (available in English and French) - (WMO-TD No. 342)
- WCDP-10 CALCULATION OF MONTHLY AND ANNUAL 30-YEAR STANDARD NORMALS, March 1989 (prepared by a meeting of experts, Washington DC, USA) - (WMO-TD No. 341)

- WCDP-11 REPORT OF THE EXPERT GROUP ON GLOBAL BASELINE DATASETS, Asheville, USA, 22-26 January 1990 - (WMO-TD No. 359)
- WCDP-12 REPORT OF THE MEETING ON HISTORICAL ARCHIVAL SURVEY FOR CLIMATE HISTORY, Paris, 21-22 February 1990 - (WMO-TD No. 372)
- WCDP-13 REPORT OF THE MEETING OF EXPERTS ON CLIMATE CHANGE DETECTION PROJECT, Niagara-on-the-Lake, Canada, 26-30 November 1990 - (WMO-TD No. 418)

Note: *Following the change of the name of the World Climate Data Programme (WCDP) to World Climate Data and Monitoring Programme (WCDMP) by the Eleventh WMO Congress (May 1991), the subsequent reports in this series will be published as WCDMP reports, the numbering being continued from No. 13 (the last 'WCDP' report).*

- WCDMP-14 REPORT OF THE CCI WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 21-25 October 1991
- WCDMP-15 REPORT OF THE CCI EXPERTS MEETING ON CLIMATE CODE ADAPTATION, Geneva, 5-6 November 1991 - (WMO-TD No. 468)
- WCDMP-16 REPORT OF THE CCI EXPERTS MEETING ON TRACKING AND TRANSMISSION OF CLIMATE SYSTEM MONITORING INFORMATION, Geneva, 7-8 November 1991 - (WMO-TD No. 465)
- WCDMP-17 REPORT OF THE FIRST SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 19-20 November 1991 (also appears as WCASP-18) - (WMO-TD No. 475)
- WCDMP-18 CCI WORKING GROUP ON CLIMATE DATA, Geneva, 11-15 November 1991 (WMO-TD No. 488)
- WCDMP-19 REPORT OF THE SECOND CLICOM EXPERTS MEETING, Washington DC, 18-22 May 1992 - (WMO-TD No. 511)
- WCDMP-20 REPORT ON THE INFORMAL PLANNING MEETING ON STATISTICAL PROCEDURES FOR CLIMATE CHANGE DETECTION, Toronto, 25 June, 1992 (WMO-TD No. 498)
- WCDMP-21 FINAL REPORT OF THE CCI WORKING GROUP ON CLIMATE DATA AND ITS RAPPORTEURS, November 1992 - (WMO-TD No. 523)
- WCDMP-22 REPORT OF THE SECOND SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 16-17 November 1992 (also appears as WCASP-22) - (WMO-TD No. 529)
- WCDMP-23 REPORT OF THE EXPERTS MEETING ON REFERENCE CLIMATOLOGICAL STATIONS (RCS) AND NATIONAL CLIMATE DATA CATALOGUES (NCC), Offenbach am Main, Germany, 25-27 August 1992 - (WMO-TD No. 535)
- WCDMP-24 REPORT OF THE TENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Geneva, 20-22 September 1995 (also appears as WCASP-34) - (WMO-TD No. 711)

- WCDMP-25 REPORT OF THE FIFTH SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 26 September 1995 (also appears as WCASP-35) - (WMO-TD No. 712)
- WCDMP-26 REPORT ON THE STATUS OF THE ARCHIVAL CLIMATE HISTORY SURVEY (ARCHISS) PROJECT, October 1996 (prepared by Mr M. Baker) - (WMO-TD No. 776)
- WCDMP-27 SUMMARY REPORT OF THE MEETING OF THE THIRD SESSION OF THE CCI WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 26 February - 1 March 1996 - (WMO-TD No. 818)
- WCDMP-28 SUMMARY NOTES AND RECOMMENDATIONS FOR CCI-XII FROM MEETINGS CONVENED TO PREPARE FOR PUBLISHING THE FIFTH AND SIXTH GLOBAL CLIMATE SYSTEM REVIEWS AND FOR A PUBLICATION ON THE CLIMATE OF THE 20TH CENTURY, July 1997 - (WMO-TD No. 830)
- WCDMP-29 CLIMATE CHANGE DETECTION REPORT - REPORTS FOR CCI-XII FROM RAPORTEURS THAT RELATE TO CLIMATE CHANGE DETECTION, July 1997 (WMO-TD No. 831)
- WCDMP-30 SUMMARY NOTES AND RECOMMENDATIONS ASSEMBLED FOR CCI-XII FROM RECENT ACTIVITIES CONCERNING CLIMATE DATA MANAGEMENT, July 1997 (WMO-TD No. 832)
- WCDMP-31 REPORTS FOR CCI-XII FROM RAPORTEURS THAT RELATE TO CLIMATE DATA MANAGEMENT, July 1997 - (WMO-TD No. 833)
- WCDMP-32 PROGRESS REPORTS TO CCI ON STATISTICAL METHODS, July 1997 (prepared by Mr Christian-Dietrich Schönwiese) (WMO-TD No 834)
- WCDMP-33 MEETING OF THE CCI WORKING GROUP ON CLIMATE DATA, Geneva, 30 January - 3 February 1995 - (WMO-TD No. 841)
- WCDMP-34 EXPERT MEETING TO REVIEW AND ASSESS THE ORACLE-BASED PROTOTYPE FOR FUTURE CLIMATE DATABASE MANAGEMENT SYSTEM (CDBMS), Toulouse, France, 12-16 May 1997 - (WMO-TD No. 902)
- WCDMP-35 REPORT OF THE ELEVENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Mauritius, 9-14 February 1998 (also appears as WCASP-47) - (WMO-TD No. 895)
- WCDMP-36 REPORT OF THE MEETING OF THE CCI TASK TEAM ON CLIMATE ASPECTS OF RESOLUTION 40, Geneva, Switzerland, 10-11 June 1998 - (WMO-TD No. 925)
- WCDMP-37 REPORT OF THE MEETING OF THE JOINT CCI/CLIVAR TASK GROUP ON CLIMATE INDICES, Bracknell, UK, 2-4 September 1998 - (WMO-TD No. 930)
- WCDMP-38 REPORT OF THE MEETING OF THE WMO COMMISSION FOR CLIMATOLOGY (CCI) TASK GROUP ON A FUTURE WMO CLIMATE DATABASE MANAGEMENT SYSTEM (CDMS), Ostrava, Czech Republic, 10-13 November 1998 and FOLLOW-UP WORKSHOP TO THE WMO CCI TASK GROUP MEETING ON A FUTURE WMO CDMS, Toulouse, France, 30 March-1 April 1999 - (WMO-TD No. 932)

- WCDMP-39 REPORT OF THE MEETING OF THE CCI WORKING GROUP ON CLIMATE DATA, Geneva, Switzerland, 30 November-4 December 1998 - (WMO-TD No. 970)
- WCDMP-40 REPORT OF THE MEETING ON CLIMATE STATISTICS, PRODUCT DEVELOPMENT AND DATA EXCHANGE FOCUSING ON CLICOM 3.1, Geneva, 25-29 January 1999 - (WMO-TD No. 971)
- WCDMP-41 PROCEEDINGS OF THE SECOND SEMINAR FOR HOMOGENIZATION OF SURFACE CLIMATOLOGICAL DATA, Budapest, Hungary, 9-13 November 1998 (WMO-TD No. 962)
- WCDMP-42 REPORT OF THE MEETING OF EXPERTS ON THE CLIMATE OF THE 20TH CENTURY, Geneva, 26-30 April 1999 - (WMO-TD No. 972)
- WCDMP-43 REPORT OF THE TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT FOCUSING ON CLICOM/CLIPS DEVELOPMENT AND EVALUATION, Niamey, Niger, 03 May-10 July 1999, (WMO-TD No. 973)
- WCDMP-44 REPRESENTATIVENESS, DATA GAPS AND UNCERTAINTIES IN CLIMATE OBSERVATIONS, Invited Scientific Lecture given by Chris Folland to the WMO Thirteenth Congress, Geneva, 21 May 1999 - (WMO-TD No. 977)
- WCDMP-45 WORLD CLIMATE PROGRAMME - WATER, DETECTING TREND AND OTHER CHANGES IN HYDROLOGICAL DATA, Zbigniew W. Kundzewicz and Alice Robson (Editors) - (WMO-TD No. 1013)
- WCDMP-46 MEETING OF THE WMO CCI TASK GROUP ON FUTURE WMO CLIMATE DATABASE MANAGEMENT SYSTEMS (CDMSs), Geneva, 3-5 May 2000 (WMO-TD No. 1025)
- WCDMP-47 REPORT ON THE ACTIVITIES OF THE WORKING GROUP ON CLIMATE CHANGE DETECTION AND RELATED RAPORTEURS, 1998-2001 (May 2001, updated from March 2001) (WMO-TD No. 1071)
- WCDMP-48 REPORT OF THE FIRST SESSION OF THE MANAGEMENT GROUP OF THE COMMISSION FOR CLIMATOLOGY (Berlin, Germany, 5-8 March 2002) (also appears as WCASP-55) (WMO-TD No. 1110)
- WCDMP-49 1. REPORT ON THE CLICOM-DARE WORKSHOP (San José, Costa Rica, 17-28 July 2000); 2. REPORT OF THE INTERNATIONAL DATA RESCUE MEETING (Geneva, 11-13 September 2001) (WMO-TD No. 1128)
- WCDMP-50 REPORT OF THE CLIMATE DATABASE MANAGEMENT SYSTEMS EVALUATION WORKSHOP (Geneva, 11-13 September 2001) (WMO-TD No. 1130)
- WCDMP-51 SUMMARY REPORT OF THE EXPERT MEETING FOR THE PREPARATION OF THE SEVENTH GLOBAL CLIMATE SYSTEM REVIEW (7GCSR) (Geneva, 16-19 September 2002) (WMO-TD No. 1131)
- WCDMP-52 GUIDELINES ON CLIMATE OBSERVATION NETWORKS AND SYSTEMS (WMO-TD No. 1185)

- WCDMP-53 GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION (WMO-TD No. 1186)
- WCDMP-54 REPORT OF THE CCI/CLIVAR EXPERT TEAM ON CLIMATE CHANGE DETECTION, MONITORING AND INDICES (ETCCDMI) (Norwich, UK, 24-26 November 2003) (WMO-TD No. 1205)
- WCDMP-55 GUIDELINES ON CLIMATE DATA RESCUE (WMO-TD No. 1210)
- WCDMP-56 FOURTH SEMINAR FOR HOMOGENIZATION AND QUALITY CONTROL IN CLIMATOLOGICAL DATABASES (Budapest, Hungary, 6-10 October 2003) (WMO-TD No. 1236)
- WCDMP-57 REPORT OF THE RA V DATA MANAGEMENT WORKSHOP (Melbourne, Australia, 28 November-3 December 2004) (WMO-TD No. 1263)
- WCDMP-58 GUIDELINES ON CLIMATE WATCHES (WMO-TD No. 1269)
- WCDMP-59 REPORT OF THE MEETING OF THE RA I WORKING GROUP ON CLIMATE MATTERS (Dakar, Senegal, 22 – 24 February 2006) (WMO-TD No. 1351)
- WCDMP-60 GUIDELINES ON CLIMATE DATA MANAGEMENT (WMO-TD No. 1376)