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

Within the World Climate Data and Monitoring Programme, a meeting was held at the kind invitation of Spain (Malaga, 24-26 February 2003) in which a number of experts in the two CCI Open Programme Area Groups (OPAGs) on Climate Data and Monitoring initiated the preparation of guidelines on metadata and data homogenization, observation networks and systems, and data rescue. The participants were either members of an Expert Team of CCI, or were invited experts.

The Guidelines on the Data Rescue, Preservation and Digitization of Climate Data are meant to be easy to read and refer to, well illustrated, and not bulky. They provide information and assistance on how to organize and implement climate services, and present processes and technological solutions that attempt to address the special situation and needs of smaller NMSs which have limited resources.

The review of the Guidelines was the first such activity that was done within the CCI OPAG structure, so that all CCI Members were given an opportunity to review and comment, as well as to see the progress being made in the OPAGs. It was drafted by a sub-group of the CCI Expert Team on the Data Rescue, Preservation and Digitization of Climate Data, and circulated for contributions and comment among the members of the CCI Expert Team.

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 NMSs with respect to size and stage of technological development along with the variability of weather types and climate, some practices may not have significant utility for specific Member. However, this document does cover a wide range of guidance that should provide some form of assistance to every Member.
Guidelines on Climate Data Rescue

L.S. Tan¹, S. Burton², R. Crouthamel³, A. van Engelen⁴, R. Hutchinson⁵, L. Nicodemus⁶, T.C. Peterson⁶, F. Rahimzadeh⁷

1. Introduction

Apart from having a good observation network for climatology, steps must be taken to make sure the vast amount of climate data collected are properly preserved in an easily accessible useful form. At an International Data Rescue Meeting among the Experts in WMO from September 11-13, 2001 in Geneva, the following definition of data rescue was reached:

Data Rescue is the ongoing process of 1. preserving all data at risk of being lost due to deterioration of the medium and; 2. digitizing current and past data into computer compatible form for easy access.

The re-establishment of the data rescue project as a high priority within WMO Programs is critically important to ensure future generations of scientists and other data users have access to all the information necessary for their studies and projects. These rescued data combined with already available data will enable better assessments of projections of the climate into the future that can serve as input for the policy makers to mitigate loss due to natural disasters and will provide increased information for economic development.

Critical Needs: Many of the world’s climate datasets contain digital data back to the 1940’s, but few have a great volume of data before this time. It will be very useful if the data before 1940 were made available and accessible.

2. Rationale

Why should we take such great trouble to rescue the historical data at risk of being lost? Recent IPCC statement on climate change indicates that to detect such change in the climate system and global warming, we should study the natural variability of the undisturbed climate with long period records of quality climate observations. To help make the agrometeorological and aeronautical models more credible, good quality, long records of climate data are also required. Most countries have conscientiously taken observations of the weather, recorded them either manually or automatically, transcribed them onto preset paper forms and eventually entered them into some form of computer media for easy access and analysis. Unfortunately, some of these precious and invaluable data have not been transcribed from the paper forms due to the lack of funds and sufficient personnel. Most of the paper forms are now at risk of being lost due to rapid deterioration of the medium. It is the purpose of this guide to provide advice, a service that can recognize, rescue, preserve and digitize such climate data. However, it cannot be stressed enough that in addition to the historical data, all current data must be made available in digital form as well.

¹ Malaysian Meteorological Service, Malaysia
² Caribbean Institute for Meteorology and Hydrology, Barbados
³ National Weather Service, NOAA, USA
⁴ KNMI, Netherlands
⁵ National Climate Centre, Bureau of Meteorology, Australia
⁶ National Climatic Data Center, NOAA, USA
⁷ Islamic Republic of Iran Meteorological Organization, Iran
When should data rescue, preservation and digitization of climate data start? The answer almost always is that it should have started before now. It is likely to also remain a continuing issue as digital data need to be transferred to new media to maintain readability into the future. A full process may not be able to be started instantly, but a small beginning should be able to be started almost immediately.

Who should undertake data preservation, rescue and digitization? First and foremost it is any group or any one who has data whether paper, microfilm, or digital. There should be cooperation between the groups. Beyond this, those who use climate data should have a special role in data rescue as users are in better positions to appreciate and value the data being rescued and they know what data are most important. These users are often in the climate sections of NHMS. But they are not limited to those official groups as users and keepers of data can be found in many institutions both public (Agricultural Departments) and private (universities, plantations, agribusiness).

Where should data preservation, rescue and digitization take place? There are two possible approaches. The first is within the country where the data to be rescued are stored. This may be in the NHMS headquarters, or in various provincial offices. The best approach may be to collect the data to be rescued into one location or, if the expertise exists or can be developed in each region, to have the imaging and preservation take place in provincial offices. Another possibility is in regional multi-national climate centers such as ACMAD. Expertise can be collected in these offices that exceed what any one nation can afford to develop. Which option is best depends on the individual circumstances of that country or region.

What should be the priority when dealing with data rescue? The International Data Rescue Meeting (Geneva, 11-13 September 2001) established the following priorities for the Data Rescue project that:

I. High quality historical climate data of importance nationally, regionally or globally that is at risk of loss or assists in completing established databases.

II. High quality current climate data of importance nationally, regionally or globally that enhances or completes the established databases.

Within the current data or historical data, the priority can be further set to be:

i. Instrumental data to be decided by usefulness into:
   a. monthly means
b. daily data
c. hourly, synoptic and high resolution data from automated weather stations
d. autographic charts (including sunshine cards)

ii. Metadata: description and definition of data (these must be rescued and digitized together with item (i)).

iii. Non-instrumental data such as visual weather observations

iv. Phenological data

Although these are the general guidelines to the priority of data, each country or centre must decide on the priority based on the prime needs of the country and the state of condition of the climate data records. The most critical data at risk of being lost must be dealt with first.

3. How to undertake the data rescue?

3.1 Search and locate: It is obvious that the first task is to know where are the data. Of paramount importance is that, before the onset of the preserving procedure, historical datasets, hidden in national archives, services, private collections etc., are identified and recognized as meaningful climatological data. If we ignore this aspect numerous potential valuable datasets can get lost without anyone knowing it! A search in the following places/sources may reveal that valuable climate data are often left forgotten apart from those already digitized and well managed in any existing data management systems in the organization or NHMS.

i. Meteorological Services

ii. Hydrological Services

iii. Other sources of data, other departments or agencies often have data holdings

iv. Universities

v. Individuals like long-serving staff in meteorological and related organizations; heads of international data centres; local historians.

The search process should cross-check with available digital records to check whether the records have already been digitized, for identifying significant gaps and for providing clues to the dates on which observing programs may have begun in particular regions.

3.2 Inventorize: Immediately after some climate data are found, they should be inventorized (take stock) to determine what data exist and what rescue efforts had been done before. Pay special attention to the inventory of related metadata.
After the search:

- Assemble all records and metadata in one central location
- Organize records according to a logical imaging plan
  - Station/year
  - Year/station
  - Form type/station/year
- Estimate image volume, image work plan.
- Format Identification: Review and identify different types
  - Hourly
  - Daily
  - Monthly
- Detailed Accounting
  - Count number of pages to be imaged
  - Define periods of record
  - Define missing documents
  - Cross check for number of images on CDs or other storage media
- Methods employed
  - Set-up an Excel (or other) spreadsheet for each station
    - By year/month/day/obs time
  - Enter number of pages to be imaged
  - Keep a log book of the metadata. Often crucial metadata are hidden in between the historical records.

Figure 3. Example of a spreadsheet containing inventory of climate records
Use spreadsheets to count the number of pages to be imaged and to keep track of the image filenames. When you image records you will find that you accumulate a large number of images whose filenames are generated by the camera software and which have no bearing to the actual date of the record or the content of the record. Some cameras may repeat filenames, if images are downloaded from the PC in the same day. So keeping track of the image names is very important, if you want to locate the records later for viewing or keying. It also helps in organizing and assembling the data records into convenient work groups for efficiently managing the image workflow.

3.3 **Preservation and storage**

**Section (a): If the data are already digital**

Data rescue is still important for data that are already electronic as the media on which they reside may not be permanent. For example, magnetic tapes lose their magnetism over time, and this process is faster in humid or warm conditions. Also, as computer technology evolves it can become hard to find a computer that can read old media. For example, do you still have a computer handy that can read the old 5½ inch floppy disks? Will you have one in 5 years? The same obsolescence problem is likely to occur in all electronic media eventually. Therefore, it is important to continually migrate the data to new or current media. If this is done often enough, it rescues the data before there is a problem and any data are lost.

**Refreshing the media**

For computer tapes, the magnetic tapes should be replaced with new ones at least once in every ten years.

As a general rule, optical discs and compact discs (CD) last longer than tapes and cartridges. As technology stands, the most hopeful form of all media and in terms of most widely used is CD and its variant DVD. As technology develops, and it is gathering pace as time goes on, it is highly recommended that the precious climate data migrate from technology to technology. When the next most reliable media appear, we should move them again to ensure their preservation. However, it must be cautioned here too that the software used to record and store the climate data must be migrated too. Flat files or just plain ASCII data files may also be kept. (For the sake of migration it is also important that the digital code in which the files are stored is very straightforward, i.e. it does not need decoding. The plain ASCII format is widely recognized and accepted to best fulfill this demand.

The following items are recommended for the case of digital forms

i. Safeguard the digital files. These images may now be considered the official records and therefore should be protected as such.

ii. Organize for future access so that users can have ready access to the data.

iii. Store on appropriate media so the data will continue to be available into the indefinite future with appropriate migration with technological advance.

iv. Off site storage of electronic copies will preserve the electronic data even in the event of a fire or flood in the main office. To safeguard the data even in the event of war may require also storing these data in a multinational regional climate center.

**Section (b): If the data are on paper or microfilm/microfiche**

The goal is to take these data and eventually scan or key-enter them to make the data useable in a digital form.

It is recommended that the following items should be considered when preserving critical data that are in danger of being lost.
a. Preserve, protect from insects, humidity, etc.
b. Don’t throw anything away even after they are scanned if resources for storage permit.
c. Microfilms are not permanent archives as they deteriorate as well. (These need special air-conditioned rooms with dehumidifiers for preservation).
d. Disperse the records. The data may be processed in a multinational center but then copies of the data should be dispersed to the ORIGINATING countries as well as being maintained in the Center.
e. Special rooms to store the climatological and hydrological documents can be created where temperature and humidity are carefully controlled and the data are protected from insects, rodents, mildew, fire, theft and all other dangers. A professional archivist could help guide this process.
f. The goal of rescuing climate observations is to provide those data in secure media and in an easy to use form. In this case, the digital forms preservation efforts, like the digital data rescue mentioned earlier, should include continual migration to new computer media. Serial data should be made available to scientific users by direct digitization of the data.

The first step is often scanning the data or recording a digital picture, which can then be preserved from further deterioration and made widely available. These objectives will be achieved through the following initiatives.

i. Electronic imaging of all manuscripts of original meteorological data via an optical scanner or digital camera

Digital cameras are much more efficient to use and it is they be used to image scanned. Scanners are cumbersome to use, expensive scanner have memories and can before having to be PC. They also come with automatically generates image.

![Digital camera](image-url)

Figure 4. A digital camera

ii. Recommended basic procedures

Imaging practices and techniques may vary considerably. However, there are some basic procedures which are strongly recommended:

a. Design spreadsheets to record information about each document or image. There will be a large number of images obtained during record imaging. Accounting and file tracking needs to be done.
b. Count all records manually before imaging begins, so that you know the total number of pages to be imaged. Log these numbers into the spreadsheets.
c. Compare the number of pages counted to the number of images actually produced and stored on CD or disk. The number of pages counted and entered into the spreadsheets should equal the number of images produced.
d. Each time that picture files from a digital camera are downloaded to the computer, the camera software generates unique filenames for each image stored on the hard disk. These filenames bear no relation to the time frame or content of information contained on the document. Log these filenames into spreadsheets, so that images can be located and retrieved by their IDs, time frame and content (station name, number, date, observation time, form type, page number, etc.)

Figure 5. Example of listing of imaged files

3.4 Validating the imaged files

First, the imaged files must be viewed and checked before being written onto a CD.

- Filenames of image documentation must be entered into a spreadsheet
- ‘Burning’ and labeling CDs together with a copy of the spreadsheet
- Making duplicate CDs
  - Working copies
  - Archival copy
3.5 **Key entry of the climate data**

Once the documents are properly imaged, the next step is digitization or key entry of the data. This is probably the most tedious and difficult process of data rescue.

These actions would be done completely based on scientific, technical and social conditions. For example, limited resources may mean that only selected data will be able to be key entered. Deciding which data are selected to be keyed should be based on scientific considerations. (Example: GCOS 2nd Adequacy report may be used to help identify global scale priorities for data rescue).

![Figure 6. Entering data](image)

When digitizing, it is important that the transcription is 1 to 1. That means that one should avoid any form of coding. If you are urged to apply coding then take care that it is well documented and reversible in the sense that one can return to the original data). Since it is NOT always possible to capture all the information digitally, it is important to retain originals or copies of originals.

In some cases, if the NMHS cannot provide the staff and funding to complete key entry of all the climate data, assistance can be sought from WMO secretariat. Through a Climate Database Modernization Program (CDMP), the National Climatic Data Center of NOAA, USA, helped many African countries and Vietnam to document images and digitize their climate data.

3.6 **Quality Checking (QC) of the climate data**

The climate data thus rescued will not be of any use unless proper quality checking and verification process have been performed on the climate data. (Ref. Guide to Climatological Practices).

The quality control (QC) of the newly key-entered digital data should be treated the same way as data from other sources. The same is true about homogeneity assessments – they should be treated the same as data from any other source. However, all aspects of the data rescue process are in need of their own unique quality assurance tests. For example, some of the digital photos may be out of focus and unreadable, which is one reason to preserve paper records. One effective approach to quality assurance is to periodically (e.g. every 300 pages) check a few images to make sure they are readable. It is also appropriate to check a few pages worth of images and compare them to the paper archive to make sure no pages of data were missed.

There are additional quality assurance steps that should be taken during and after the project. A prime example of during production QC – undertaken while the work is going on - is to compare the number of images or months of digital data created with the number expected from the count of pages to be scanned or the period of record for the station. At the beginning of this work, an inventory was made of available data. At the end of the work the quantity of digital data created should be compared to the initial inventory to see if there is any indication that data were missed. And lastly, were any insights gained during this process that might lead to locating other sources of data in need of being rescued?

QC of historical data can be more challenging since there is less neighbouring data available for the older data.
3.7 **Analyses and climate products**

With the help of the metadata the data series should be analyzed, as soon as possible in the digitizing process, on their meaning and reliability. Next, if possible, the data series should be tested with standardized homogeneity tests, preferably on the basis of monthly aggregations.

Once the data rescued are found to be of good quality and integrity, standard climate products such as tables, climate averages and graphs should be made such as those mentioned in the Guide to Climatological Practices. This not only makes good use of the data but is a way to make sure that the data are in good shape and lets other groups (such as universities) know that these newly rescued data are available for their use as well.

3.8 **Make data available to users**

The last task in data rescue will be to make the climate data available to users in the most convenient accessible form either through the internet or other traditional means such as publications, diskettes or CDs.

3.9 **Linkage to the management**

The data rescue process and strategy must be linked to the data management policies of the institution and make sure they fit in properly.

4. **Possible obstacles**

Knowing how to rescue data in general does not mean that the process will be simple or easy. Some obstacles exist. These include:

There will be new equipment for the individuals who will undertake the data rescuing work to use (scanners, digital cameras, data bases for digitizing data, etc.) in which they will need training and experience. The whole process of rescuing data can be expensive, which is one of the reasons it may not have been done yet. This may require some outside support or it may simply require making a clear case for the need so that there is determination to make sure this type of activity is supported by the NMHS. Another resource that should be lined up before undertaking data rescue is other individuals and groups who have already done this work. Not only does consultation with them provide additional information, but it also builds a focus group to consult with if problems occur along the way. NMHS having problems may contact the WMO Secretariat, so that aid from other Members can be made available.

5. **Summary and conclusion**

A lot of work, but very worthwhile. You are not alone as NMHS all over the world are doing this too. Good luck.
GLOSSARY

ACMAD: African Centre of Meteorological Applications for Development

ARCHISS: WMO/UNEP Archival Climate History Survey Project

ASCII: American Standard Code for Information Inter exchange is the most common format for text files in computers

Autographic charts: Charts recorded by a clock-worked drum automatically producing meteorological measurements such as temperature, wind direction and speed and atmospheric pressure

“Burning” CDs: A process to copy computer files to compact discs

Digital Camera: A digital camera mounted on a stand to create a digital image file of the original analogue records

CDMP: Climate Database Modernization Programme of the National Oceanic and Atmospheric Administration, USA

CLICOM: CLimate COMputing is a project of the World Climate Data and Monitoring Programme (WCDMP), whose task is to transfer methods and techniques for processing climatological data and assist users

Computerization: Using computers in the processing of data.

DARE: An acronym for Data Rescue and used for the WMO Project on climate data rescue.

Digital data: Data that had been key-entered into computer and stored as numbers as against analogue or charts. These numbers can either be in plain text forms or organized into a relational database such as Microsoft ACCESS or ORACLE

Digitization: A process to transcribe analogue data into digital form for processing by a computer

GCOS: Global Climate Observing System (IOC/WMO/ISSU)

Image: Here we refer to a picture form of the climate data document captured either by a scanner or a digital camera. The smart cards or other computer media can store the images

IPCC: Intergovernmental Panel on Climate Change

Metadata: Is a set of attributes, or elements, necessary to describe a resource. In data rescue, these refer to both station information and climate data inventory information

Microfilm: A method of reproducing images of records at greatly reduced size on photographic film

Microfiche: Rectangular sheet form of microfilm

Migration: A means of overcoming technological obsolescence in hardware and software by moving data from one computer medium to another type of computer medium to enable the preservation of the intellectual content of the digital object

NMHS: National Meteorological and Hydrological Services
**Phenological data:** Data that describe the cyclic and seasonal natural phenomena, especially in relation to climate and animal life

**Preservation:** Keeping the data safe from harm or danger

**Reformatting:** Copying information from one storage medium to another or converting from one file format to another

**Refreshing:** Copying information from one storage media to the same storage media

**Scan:** To produce an image file from an analogue record or hard copy of a digital record with a scanner

**Scanner:** A device used to scan documents into a digital file format

**Spreadsheet:** A type of format used especially in accounting. Here we refer to the spreadsheet software either by Microsoft EXCEL or Lotus 1-2-3

**References and resources**

WCDMP-No. 49 Reports on the CLICOM-DARE Workshop, and on the International Data Rescue Meeting, Geneva, Switzerland, 11-13 September 2001, WMO-TD No. 1128

The second edition of the Guide to Climatological Practices

The completed sections of the draft of the Third edition of the Guide to Climatological Practices

The second report on the adequacy of the global observing system for climate in support of the UNFCCC (Draft report version 4.1 (20.12.02)

WCDMP-No. 26 Report on the Status of the Archival Climate History Survey (ARCHISS) Project, October 1996 (prepared by Mr M. Baker) - (WMO-TD No. 776)


Donovan, E. 2000, *Microfilm Standards*, E-mail, National Archives of Australia

WCDMP-No. 30 Summary notes and recommendations assembled for CCI-XII from recent activities concerning climate data management, July 1997 (WMO-TD No. 832)
WCDMP-No. 31 Reports CCI-XII from Rapporteurs that relate to climate data management, July (WMO-TD NO. 833)

WCDMP-No. 48 Report of the first session of the CCI Management Group of the Commission for Climatology (Berlin, Germany, 5-8 March 2002) also appears as WCASP-55, WMO-TD No. 1110