Guidelines on Best Practices for Climate Data Rescue

2016 edition

WEATHER CLIMATE WATER

WORLD
METEOROLOGICAL
ORGANIZATION

WMO-No. 1182

Guidelines on Best Practices for Climate Data Rescue

2016 edition

WORLD
METEOROLOGICAL
ORGANIZATION

WMO-No. 1182

NOTES

METEOTERM, the WMO terminology database, may be consulted at http://public.wmo.int/en/resources/meteoterm.

Readers who copy hyperlinks by selecting them in the text should be aware that there will be additional spaces immediately following http://, https://, ftp://, mailto:, and after slashes (/), dashes (-) and unbroken sequences of characters (letters and numbers). These spaces should be removed from the pasted URL. The correct URL is displayed when hovering over the link or when clicking on the link and then copying it from the browser.

Complementary information on best practices for data rescue is provided in the International Data Rescue portal I-DARE (www.idare-portal.org), including illustrations and photos, more detailed practice and equipment specifications, data rescue success stories, weblinks, contact information, etc. Any additional information needs should be addressed through I-DARE.

ACKNOWLEDGEMENT

We would like to thank Ms Nancy Westcott, retired from Midwestern Regional Climate Center, Illinois State Water Survey of University of Illinois, United States of America, for her outstanding contribution to this publication.

WMO-No. 1182

Chairperson, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0) 22 730 84 03
Fax: +41 (0) 22 730 81 17
E-mail: publications@wmo.int

ISBN 978-92-63-11182-1

CONTENTS

Page

SCOPE v

1. INTRODUCTION – THE SCIENTIFIC IMPORTANCE OF CLIMATE DATA RESCUE 1

2. OVERVIEW OF CLIMATE DATA RESCUE: ARCHIVING, IMAGING AND DIGITIZING . . 1

3. ARCHIVING PAPER MEDIA 2
   - 3.1 Searching and locating 2
   - 3.2 Preservation and storage 3
   - 3.3 Creating electronic inventories of paper/microfilm holdings and images 4

4. IMAGING ORIGINAL MEDIA 5
   - 4.1 Basic steps in the imaging process 6
   - 4.2 Validation and storage of image files 6

5. DIGITIZING DATA VALUES 7
   - 5.1 Minimizing keying errors 8
   - 5.2 Comparison with international databases 8

6. ARCHIVING DIGITAL MEDIA (IMAGES AND DATA VALUES) 9
   - 6.1 Cross-checking printed media, images and digital datasets 9
   - 6.2 Refreshing the media 9
   - 6.3 Recommendations for archiving digital media 9

APPENDIX 1. A GLOBAL ACTIVITY 11

APPENDIX 2. DATA RESCUE ASSESSMENT 12

APPENDIX 3. INFRASTRUCTURE, EQUIPMENT, SUPPLIES AND PERSONNEL 13

APPENDIX 4. GUIDANCE ON MAINTAINING PAPER ARCHIVES 14

APPENDIX 5. PRIORITIZING CONSIDERATIONS FOR IMAGING AND DIGITIZING DATA . 15

APPENDIX 6. ELECTRONIC IMAGING TECHNIQUES 17

APPENDIX 7. OTHER DIGITIZATION METHODS 20

APPENDIX 8. DIGITIZATION ERRORS AND QUALITY ASSURANCE PROCESS 22

APPENDIX 9. LESSONS LEARNED 23

APPENDIX 10. APPLICATIONS TO SHOW VALUE OF DATA 25

APPENDIX 11. CHECKLIST FOR CLIMATE DATA RESCUE 26

APPENDIX 12. GLOSSARY 28

APPENDIX 13. REFERENCES 30

SCOPE

This technical document is an update of WMO/TD-1210, WCDMP-55, *Guidelines on Climate Data Rescue* (2004). It builds on the original Guidelines, while taking into account both changes in technology that have occurred in the intervening 12 years and lessons learned in more recent climate data rescue activities around the world. An overview of data rescue is presented with chapters on its importance, archiving original media, imaging, digitization and archiving digital images and digital data. Twelve appendices provide supporting information.

The Guidelines on Climate Data Rescue are intended to provide guidance in the form of recommended best practices. Because of the diversity of National Meteorological and Hydrological Services (NMHS) with respect to the size and stage of technological development, along with the variability of weather types and climate, some practices may not be useful for every WMO Member. That being said, the Guidelines cover a wide range of guidance that should provide assistance on how to organize and implement data rescue and provide generalized technological solutions for every Member. More specific technological information, as well as informative illustrations and photos, may be found at the International Data Rescue (I-DARE) portal that is maintained by WMO with the assistance of the Royal Netherlands Meteorological Institute and the WMO Commission of Climatology Expert Team on Data Rescue.

While specific to weather and climate data, these best practices could also be applied to the rescue of data in other scientific fields, both within the remit of WMO and beyond. In particular, the rescue of hydrological, marine and other environmental data follows similar overall principles and practices and is basically considered to be within the scope of these Guidelines. Specificities of such data, however, need to be identified and taken into account in close collaboration with the respective communities, including, for example, the WMO Commission for Hydrology and the WMO–Intergovernmental Oceanographic Commission (UNESCO) Joint Technical Commission for Oceanography and Marine Meteorology.

GUIDELINES ON BEST PRACTICES FOR CLIMATE DATA RESCUE

1. INTRODUCTION – THE SCIENTIFIC IMPORTANCE OF CLIMATE DATA RESCUE

Climate data rescue involves organizing and preserving recorded instrument observations and climate data at risk of being lost. Data rescue is critically important to ensure that future generations of scientists and other data users have access to all the information necessary for assessing climate variability and change, as well as providing a range of climate services. These data help bridge the gap between palaeoclimate data and current observations. Many of the world's climate datasets contain digital data back to the 1940s and 1950s, but many observations from the 19th century and early 20th century exist only in hard-copy forms and need to be rescued and digitized. Some areas of the world are well represented back to the 1800s, but many are not. The validity of climate models and palaeoclimate data (e.g. tree ring, ice core, pollen) observations would greatly benefit from long instrumental time series in all regions of the world.

Data-sparse regions remain in the 21st century. It is thus crucial to obtain and verify the more recent records. The reliability of meteorological observations can be greatly influenced by station location and siting and the quality control (QC) of observations. Comparison of station data with those from nearby sites and the ability to view original observation documentation (and original information on observational practices) will allow climatologists to better verify past and present weather and climate regimes.

Extending the climate record backwards in time and filling in temporal and spatial gaps has a number of benefits:

(a) It helps make agrometeorological, disease vectorization and hydrological and climatological numerical model representations more credible and enables better projections of future climate;

(b) Combining rescued data with data already available allows current weather and climate to be better placed within a historical perspective;

(c) It provides a basis to assess historical sensitivities of natural and man-made systems to climate variability and change, thereby better enabling accurate assessments of the impacts of future climate variability and change.

Such assessments can serve as input for policymakers to mitigate loss from weather disasters and provide increased information for economic development.

In order to be accessible to climate service providers and researchers in any practical sense, the data must be available in digital form (both images and numerical values). Furthermore, all current data must also be preserved and made available in digital form, since these observations are the historical data of the future. Original documentation is crucial for verification of digital values.

2. OVERVIEW OF CLIMATE DATA RESCUE: ARCHIVING, IMAGING AND DIGITIZING

Climate data rescue involves organizing and preserving climate data at risk of being lost due to deterioration, destruction, neglect, technical obsolescence or simple dispersion of climate data assets over time. Non-digitized data are at risk, owing to the vulnerability of the original paper record. Date rescue includes: organizing and imaging paper, microfilm and microfiche records; keying numerical and textual data and digitizing stripchart data into a usable format; and archiving the data, metadata and the quality-control outcomes and procedures. An overview of components of climate data rescue activities is provided in Table 1.

Table 1. Components of climate data rescue activities

Components of data rescue	*Activities*	*Keywords*
Archive paper and microfilm/ microfiche media (section 3)	Search and locate	NMHS, observation sites, universities, aviation and maritime agencies, agricultural organizations, international libraries and databases, national archives
	Preserve and store	Clean media; place in acid-free, labelled archive boxes, safe from dust, moisture, pests
	Create electronic paper/microfilm holdings inventory	Catalogue all paper media; estimate scope of imaging and digitizing effort
Image media (section 4)	Create master image inventory	
	Image and validate Update master image inventory Create image file inventories on each CD/DVD or in each computer directory	Update master image inventory after images are validated as readable, including metadata Cross-check CD/DVD file inventories with master image inventory
Digitize data values (section 5)	Create digital data inventory	
	Key entry, chart trace	Data input into CDMS
	Quality-control data	Update digital data inventory as data are digitized and pass through various QC tests
Archive digital media (section 6)	Cross-check printed media, images and digital data	Compare image and digital data inventories with original electronic paper/microfilm holdings inventory
	Back up electronic media	Daily
	Disperse multiple copies of images and digital data archives	To various locations
	Refresh media and migrate technologies	Every 5 to 10 years

Most countries have conscientiously taken observations of the weather, recorded them either manually or automatically, transcribed them onto paper forms or created microfilm copies and eventually entered them into some form of computer media (ideally into a Climate Data Management System (CDMS)) for easy access and analysis. Unfortunately, these data were sometimes not transcribed from the paper forms, either because of a lack of funds and/or insufficient personnel or because the data were held elsewhere, sometimes overseas. Many of the paper forms and microfilm/ microfiche copies are at risk of being lost due to rapid deterioration of the medium or because of undocumented relocation of portions of archives.

It is the purpose of these Guidelines to provide advice on the steps required to organize, image, digitize and preserve such climate data.

3. ARCHIVING PAPER MEDIA

3.1 Searching and locating

The first task of data rescue is to locate the data. Before beginning preservation procedures, it is of paramount importance that historical datasets are identified and recognized as meaningful climatological data. These may be conserved in national archives and services, at observing stations or in private collections, etc., of which the meteorological community is not always aware. Some data may even be held in overseas archives, especially those from countries with a colonial past. Without undertaking such a search, numerous potentially valuable datasets may be left unidentified and inaccessible without anyone's knowledge. Searches in the following locations and sources may reveal valuable climate data that could have been forgotten, apart from those already organized, digitized and well managed in an existing data management system within the organization or NMHS:

(a) NMHSs and other departments or agencies often have data holdings;

(b) Universities, colleges, high schools and individuals (such as long-serving staff in meteorological and related organizations);

(c) International data centres;

(d) Libraries or local historians may also have paper, microfilm/microfiche or even digital copies of data;

(e) Agricultural organizations such as large corporate food companies maintained plantations throughout the world with meteorological records dating back hundreds of years;

(f) Military agencies may also have records that may be obtainable, as might transportation ministries with aviation and maritime weather records;

(g) Religious organizations, particularly those with an observational or scientific interest, such as teaching orders, and individual clergy;

(h) Scientific societies, especially past organizations (Natural History Societies);

(i) Museums;

(j) Ships' logs, historical newspapers and personal diaries are also valuable sources of weather and climate information.

The search process should cross-check available digital records to determine whether the records have already been digitized, to identify significant gaps, and to provide clues to the dates on which observing programmes may have begun in particular regions. This includes checking international databases, such as:

(a) The Global Historical Climate Network;

(b) The International Surface Pressure Databank;

(c) The Integrated Surface Database.

In addition, national libraries contain valuable datasets (see the I-DARE portal for links to these libraries).

3.2 **Preservation and storage**

The paper and microfilm/microfiche holdings should be organized in a logical manner and stored in acid-free archive boxes on sturdy shelves or in filing cabinets:

(a) Depending on the previous location and condition of paper charts or forms, the paper copies may have to be dusted or vacuumed prior to storage;

(b) Correct work, health and safety procedures need to be followed in handling possibly contaminated records;

(c) Specialist equipment and training are required for dust removal from fragile records;

(d) Special care needs to be taken to ensure operator safety if records are mouldy or have been treated with mould inhibitors or pesticides;

(e) Fragile documents should be handled with gloves;

(f) Depending on the amount of dust, protective clothing and masks may be necessary.

Special rooms are often designated for the storage of archived paper and microfilmed data:

(a) Ideally, the rooms should be temperature- and humidity-controlled;

(b) Data must be protected from insects, rodents, mildew, fire, flooding, dust, theft and all other dangers;

(c) Avoid the presence of wood in the room to minimize the dangers of fire and insect infestation;

(d) A professional archivist could help guide this process and a manager should be assigned to periodically inspect the room.

Data are usually stored in acid-free archive boxes or in filing cabinets by type (e.g. form or chart type) and by station, year and month:

(a) The number of years of data and the number of stations may determine the most logical manner to organize the paper copies, whether by year or station;

(b) Label the archive boxes and/or filing cabinets and add the storage location (the box or file drawer number), as well as the station/year/month, media type and form type to the electronic paper/microfilm holdings inventory so that the data can be easily found for potential imaging and digitizing.

For long-term storage the following should be considered:

(a) After critical data are imaged, the original documents must not be thrown away. If resources do not permit continued storage, approach national archives, agency/service libraries, or university libraries for storage. If these institutions are unable to store archives and, if it is permitted by the national archive agency, international libraries or institutions for storage should be approached (for further discussion, see Appendix 4);

(b) Microfilm is not suitable for a permanent archive. Microfilm will deteriorate unless housed in special air-conditioned rooms with dehumidifiers;

(c) Digital images are also not a permanent archive (see section 6.2).

Appendix 3 contains further discussion of required infrastructure, supplies and personnel. Appendix 4 offers suggestions on maintenance of the paper archives.

3.3 Creating electronic inventories of paper/microfilm holdings and images

As the paper media are organized and stored, a series of inventories is required to facilitate moving from paper copies to electronic images to digital data. This is necessary so that, once the paper data are discovered and organized, they can be accessed for imaging. Once images are made, they can be found for digitizing and, once the images are digitized, both the original paper data and the electronic images can be easily retrieved for quality-control verification of digitized values. Ideally, the naming protocol at each step should conserve as much information as possible. Table 2 summarizes the steps needed to develop an archive of original media (paper and microfilm/microfiche) holdings, and to follow through with imaging, the accounting of images and the location of image files.

In summary, as climate data are found, they should be incorporated into an electronic paper/microfilm holdings inventory to determine what data exist, what rescue efforts have already been made and what remains to be done. The electronic inventory holdings include station, year, month, media type, form number or type, element type, the number of pages (including metadata pages) and, in particular, the box/file drawer location and the unique identifiers that are used to denote the documents. The variable type also should be included.

Having created an electronic inventory for a nation's historical climate records, it should be maintained and updated as new data are received and new image locations should be added to the master image inventory as imaging proceeds. The master image inventory can be used to estimate the progress of the imaging task. More details of the imaging task are discussed in Chapter 4.

4. IMAGING ORIGINAL MEDIA

After the paper and microfilmed copies are collected and catalogued, planning can begin to determine the order of imaging and digitization. Where to begin and how these tasks will proceed will be based on the needs of the individual NMHS. Prioritization will likely be different for imaging than for digitization. See Appendix 5 for a discussion of items to consider when prioritizing the order of imaging and digitizing.

Paper and/or microfilm pages are photographed or scanned to create an image or digital picture which can then be preserved from further deterioration and made widely available. All pages of original meteorological documents should be imaged via a digital camera or optical scanner. This is to ensure that all available metadata (e.g. observing practices, site and data descriptions) are rescued. See Appendix 6 for a discussion of the relative merits of digital cameras and scanners and for desirable camera features.

Table 2. Steps to develop and maintain electronic inventories

Assemble all records and metadata for storage	
Organize the records according to a logical archival plan	By station/form type/year By form type/station/year By year/station If chart data, also by meteorological element (wind, precipitation, etc.)
Create an electronic inventory of paper/microfilm holdings	Including station, year and box/file drawer location, medium type, form type Include variable types Define missing periods Estimate quantity of medium types, form types and image volume to determine a workplan
Create master imaging inventory from the electronic holdings inventory	Identify station, year, month, media type, form type or number Define periods of record Estimate number of pages to be imaged
Create image file inventories as images are created on each CD/DVD for easy location of image files or in each computer directory	
Include pages containing only metadata (information on networks, observing practices, etc.)	
Maintain the master imaging inventory as imaging proceeds	Enter number of pages imaged as images are validated Enter CD/DVD location or computer location
Cross-check the master imaging inventory with the image file inventories for number of images on CDs or other storage media	

In all cases, the metadata (the description of the data and the observing practices, location and site characteristics) must be rescued and digitized together with the images and digital values.

4.1 Basic steps in the imaging process

Imaging practices and techniques may vary, but all require documentation- and file-tracking. Some basic procedures that are strongly recommended are:

(a) As in Table 2, create a master image inventory from the electronic paper/microfilm holding inventory. Include the number of pages to be imaged. Update the master image inventory when imaging and validating the images by adding the number of pages imaged and validated and by adding the computer directory or CD/DVD location (provide template);

(b) Validate the imaged files for readability as imaging proceeds to make sure the image is readable and the image content matches the filename;

(c) A large number of image files will be created by a digital camera or scanner. Each time an image is made and downloaded to the computer, the camera/scanner software will generate a filename for each image. These filenames may not be related to the file content. Some filenames may be repeated if images are downloaded numerous times to a computer on the same day. It is recommended to employ software to generate a filename that is related to the image content (see example in (d) below). Software is available with some cameras and scanners to perform this task or can be purchased separately. See the I-DARE portal and Appendix 6;

(d) The image filename should include the station identification number (SID), an acronym for the form type (ACR), the date (YYYYmmDD) and page number (PPP). They are typically .png, .jpg or .tif files such as: SID_ACR_YYYYmmDD_PPP.png;

(e) Store images in a station/year/month directory structure. Maintain a master list of filenames included in the main directory and similar file lists in individual directories or CD/DVDs, so that images can be located and retrieved;

(f) Update master image inventory with the number of pages imaged and compare the number of pages originally counted with the number of images actually produced and stored on CD/DVD or hard drive. The number of pages counted and entered into the inventories should equal the number of images produced. Enter the location of the images (CD/DVD or directory name);

(g) Documentation of the imaging process, including the filenaming convention is crucial, as personnel involved in imaging may change over time.

4.2 Validation and storage of image files

View and check imaged files before writing them onto a CD/DVD or hard drive or into a CDMS:

(a) Images may be out of focus due to camera position;

(b) Images may be too pale and penned/pencilled corrections may not be visible;

(c) Images may be too dark because of camera/scanner settings;

(d) Portions of imaged data may be obliterated by a bright light. Use of direct lighting either with a camera flash or camera-stand lighting can reflect the light back into the camera lens, obliterating some of the imaged data, especially when the medium is glossy paper.

An effective approach to quality assurance (QA) is periodically (e.g. every 300 pages) to check a few images to make sure they are readable. It is also appropriate to check a few pages of images and compare them with the paper archive to make sure no pages of data were missed. It is important to confirm that the image filename and the image contents match at least once per month/year/station.

If CD/DVDs are used, they must be labelled.

(a) When used, the filenames of image documents must be entered into a file inventory, along with various identifying metadata (station name and number, year, month, form type, number of pages) and CD/DVD label;

(b) A copy of the file names associated with each CD/DVD should be included on that CD/DVD and in the master directory;

(c) The CD label should be included in the master image inventory.

It is of paramount importance that duplicates be created of the image archive on CD/DVD and external hard drives and/or in a CDMS. If CDs are copied to hard drives or CDMSs, each CD/DVD should have a directory and all file inventories must also be copied. The use of DVDs rather than CDs is recommended due to their greater storage capacity.

A fundamental requirement of good data management is to store backup copies of the data and metadata offsite as insurance against loss in case of damage to buildings and computers from natural or man-made disasters.

(a) It is important to distribute copies of the entire electronic image archive and the master image inventory;

(b) If data are processed in a regional or multinational centre, copies of the data should be distributed to the originating countries in addition to being maintained in the regional centre. When planning to process data at a regional centre, a prior agreement between the countries must be approved that stipulates what is to be done if natural or man-made disasters damage buildings or computers. The WMO Secretariat can assist with this type of agreement.

5. DIGITIZING DATA VALUES

Both the tracing of autographic stripcharts and the transcription of textural data into a digital form that can be stored as numbers readable by computers are referred to as digitization. The former is addressed in Appendix 7; digitization of alphanumeric records by keying is addressed in this section. Digitization is often performed at the NMHS, but sometimes also at regional centres by companies or via citizen-science initiatives and is usually guided by subject-matter experts, such as climatologists. If printed data are located within a national archive, it is likely that keying may need to be done there. When planning to process data outside the NMHS – or even outside the country – an agreement between the parties must be approved beforehand. Again, this can be done with the assistance of the WMO Secretariat.

Digitization by keying usually takes much more time than imaging, perhaps 20 hours of digitizing for one hour of imaging. Because digitizing is such a lengthy process, limited resources may prohibit all data from being digitized. The priority of data to be digitized should be based on scientific, technical and socioeconomic considerations. When it is not possible to capture all the information on a form digitally, it is especially critical to retain the original forms and copies of originals. See Appendix 5 for a further discussion on digitization priorities.

In some instances, when funds are not available to key large quantities of data, a crowdsourcing approach is used. Volunteers key the data into an electronic form or trace time-series plots from stripcharts on specially designed websites. Another approach is the use of optical character recognition (OCR) software to digitize printed material. This approach is undergoing testing and may become more effective and widely used in the future. At this time, OCR is of limited value as it requires specialized forms for readability. For a further discussion of crowdsourcing, stripcharts and OCR, see Appendix 7 and the I-DARE portal.

A first step in digitizing is to design a digital data inventory similar to the one created for imaging to follow the progress of the digitization process. Data should be organized by station, year, month and data type.

5.1 Minimizing keying errors

There are a number of techniques to minimize digitization errors. The first is to double- or triple-key the data (have two or more individuals key the same data). This can be done by having multiple keyers enter the exact same data and then compare each data value. Alternately, (a) the keyer(s) enter(s) daily data and (an)other keyer(s) enter(s) monthly data, followed by the evaluation of the consistency of the monthly total and averaged values. Double-keying can be done outside a CDMS and data imported after comparison. In this way, data can be keyed directly into the CDMS.

A second technique to reduce keying errors is to create a template of the form being keyed. Use of a template allows the keyer to know exactly where to find the data on a page, reducing the number of errors. Data keyed into a template can be directly imported into a database. Even when keying only a portion of a form, it is recommended that a keying template be created for each form type.

Thirdly, it is critical to transcribe the data exactly as they are printed (i.e. to make a 1-to-1 transcription). This is known as “key as you see” in Australia and “saisir ce que l’on voit” in France. Avoid all forms of “on the fly” coding or unit conversion. Computers are very efficient at recoding, changing units, averaging and even re-recoding. Keying as printed can ensure a definitive record of the formulae used to recode the data. If coding must be used, then take care that it is well documented and reversible (i.e. one can return to the original data). Data must be keyed exactly as they appear on the original observation form, even if it is obvious that the observer made an error. These data are occasionally used in legal cases and it is important to maintain the original data. Of course, erroneous values also must be flagged and corrected.

Double/triple keying, the creation of a template, and 1-to-1 transcription will greatly reduce errors. After the data have been keyed or imported into the CDMS as provisional data, quality control should be performed (see Appendix 8). This cannot be overemphasized.

Finally, it is recommended to key all data on a given page, in large part because it is unlikely that there will be another opportunity to return to that page and key any omitted data (e.g. latitude, longitude, observer name, observation time, other meteorological variables and all numerical and textual data). Digitizing all variables also facilitates better quality control of the record.

5.2 Comparison with international databases

Acquisition of digital climate data stored in international databases obtained from SYNOP or METAR codes may be attractive to an NMHS. Often these digital data were not stored in-house at the time they were generated and may not be easily available due to bandwidth issues required to download very large data files. These data are desirable as they are already digitized and may contain more times and variables than stored in an in-house climate database. In many instances, however, they may be plagued by keying, coding and transmission errors. They are also likely to have been subject to quality-assurance procedures. However, if a value is not clear on the original paper forms, it may be recovered from the SYNOP code. Digitization of original media and comparison with data from international databases will lead to a more complete and accurate version of climate data. Further data provenance is ensured by having the data from the international databases, the newly digitized and the images, wherever possible.

6. ARCHIVING DIGITAL MEDIA (IMAGES AND DATA VALUES)

6.1 Cross-checking printed media, images and digital datasets

While the work is underway, it is useful to record the progress of the project. This is done by cross-checking the number of months of images and digital data created with the number expected from the number of pages imaged and to be imaged, and with the period of record for each station. At the beginning of the data rescue process, an electronic paper/microfilm holdings inventory was made of original paper/microfilm data. At the end of the work, the quantity of digital data created (found in the image and digitization inventories) is compared to the initial electronic holdings inventory to determine if any data were missed.

6.2 Refreshing the media

Data rescue is just as important for digital data as paper data, since the media on which they reside are not permanent. For example, magnetic tapes lose their magnetism over time, particularly in warm or humid conditions. Paper starts to crumble after centuries, but magnetic tape media and other computer-readable media become unreadable in a matter of decades. Magnetic tapes, optical discs and 8-mm cassette tapes are generally obsolete. If digital data are still stored on these media, the national archives or universities should be contacted to move data from them to current media such as DVDs or a hard drive. Otherwise, the outcome might be an unintelligible string of numbers and characters.

As computer technology evolves, computers, computer operating systems, computer languages and the software used to read the old media also become obsolete. The problem of obsolescence will continue. It is highly recommended that, in the future, both the imaged and digital climate data be migrated to the next new, but established media (e.g. CDs, hard drives). CD/DVDs are widely used now, but may not be used in the near future. Also, it is recommended to keep multiple backups, since computers (as well as CDs and hard drives) fail catastrophically from time to time.

It must be emphasized that the software to record, store, read and interpret the climate data must be migrated from older computers and computer technology and updated as well. If this is done often enough, the data are rescued before there is a problem and data are lost. The decision to migrate media should be reviewed perhaps every two to five years and to update computer and computer software perhaps every five to seven years.

Finally, plain ASCII text data files are the preferred format for the easy reading of digital data and for moving from one computer to the next. If databases are used, the data must be easily exportable to an ASCII format (such as a .csv file). For the sake of migration, it is also important that the digital code in which the files are stored is documented and straightforward (i.e. it does not need decoding).

6.3 Recommendations for archiving digital media

(a) Safeguard both digital image and digital data files. The image files may be considered the official records and should therefore be protected;

(b) Organize the data for easy future access (so that the data can be accessed by CD or by directory structure: station, year, month). Retain an inventory of files on each CD and in the master directory;

(c) Ensure that the CDMS is backed up on a daily basis;

(d) Store digital images and data on appropriate media so the data will continue to be available into the indefinite future, with migration of datasets as required to mitigate the risks of obsolescence and decay;

(e) Use off-site storage of digital media to preserve the electronic images and digital data. To safeguard the data in the event of a man-made or natural disaster, store copies of all these data in multiple locations, including in a multinational Regional Climate Centre outside the immediate region.

APPENDIX 1. A GLOBAL ACTIVITY

1. WHO SHOULD UNDERTAKE DATA RESCUE?

Any group or individual who has data (paper, microfilm or digital) should attempt to facilitate climate data rescue. Those who are responsible for the management of a nation's climate record should have a special role in data rescue, since they are in a better position to appreciate and value the data being rescued and to know which are most important. These staff are often located in the climate sections of NMHSs and data rescue is a component of their custodian responsibilities as climate data managers. However, data rescue proponents can be found in many institutions, both public (agricultural departments) and private (universities, plantations, agribusiness). In addition, volunteer organizations, such as the International Environmental Data Rescue Organization (IEDRO) and the International Atmospheric Circulation Reconstructions over the Earth (ACRE) have been created to facilitate data rescue. Other parties interested in assisting with data rescue activities include working and retired climatologists, librarians, historians, students and spouses. Such persons would also be valuable in implementing crowdsourcing operations.

2. WHERE SHOULD DATA RESCUE ACTIVITIES TAKE PLACE?

There are a number of approaches to selecting the location for data rescue. One could be to select a location within the country where the data to be rescued are stored. This may be in the NMHS headquarters and/or in various provincial offices. The best approach may be to gather the data to be rescued in 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 to work in regional, multinational climate centres, such as the WMO Regional Climate Centres. The advantage is that expertise and equipment can be collected in these offices that exceed what individual nations may be able to afford. An alternate approach is to have a third party perform the imaging or digitizing outside the country or region, with all documents returned to the originating country. The best option depends on individual circumstances. Before the data are transferred to a third party, the WMO Secretariat would be happy to provide input on planning the location of data rescue activities.

3. THE INTERNATIONAL DATA RESCUE (I-DARE) PORTAL

The I-DARE portal is a web-based resource for people interested in data preservation, rescue and digitization. It provides a single point of entry for information on the status of past and present data rescue projects worldwide, on data that needs to be rescued and on the methods and technologies involved. It is a gateway for the exchange of information on all aspects of data rescue, including established and emerging rescue technologies. Such information also has applicability to other scientific disciplines.

Because its goals are to enhance the visibility of existing data rescue activities, to stimulate new ones and to better coordinate international data rescue efforts, the I-DARE portal is a useful communication tool. Each project listed thereon allows the excellent work of each county to be viewed by others with a similar interest. The I-DARE portal focus on informing the community on past and existing data rescue activities will also assist in identifying gaps and opportunities, help prioritize data rescue in regions where it is most needed and aid in attracting funding for projects.

APPENDIX 2. DATA RESCUE ASSESSMENT

Table 3 provides a general guide for those involved in data rescue assessments for new climate data rescue programmes.

Table 3. Checklist for data rescue assessment

1.	Review I-DARE portal for technical resources and best practices related to data rescue
2.	Download any known inventories of data from international databases, libraries, literature to assess the country's known meteorological data assets
3.	Determine the current status of data rescue in the country
4.	Visit and evaluate existing archive rooms
5.	Talk to director and staff to learn why data rescue is important to the NMHS and its data rescue priorities
6.	Advise on development of inventory of original sources, data and metadata, if necessary
7.	Advise on paper storage, space and supply needs and work, health and safety considerations
8.	Advise on methods and equipment needed for imaging
9.	Assess CDMS capabilities and infrastructure relevant to data management (age, backup, contents)
10.	Review current digitization and QA processes to see how newly digitized data can fit into that work stream and QC process or if current digitization and QC process needs revision
11.	Advise on methods and equipment needed for digitizing
12.	Advise on imaging and digitizing inventory
13.	Advise on metadata collection, archival and imaging (WMO/TD-1186)
14.	Advise on key entry of data
15.	Advise on potential use of rescued data to show value to donors/NMHSs
16.	Advise on personnel required for imaging and digitizing steps
17.	Estimate timeline of each component of data rescue based on size of task and number of staff
18.	Develop data rescue plan based on priorities of the NMHS
19.	Add data rescue mission information to I-DARE portal

APPENDIX 3. INFRASTRUCTURE, EQUIPMENT, SUPPLIES AND PERSONNEL

In all cases, storage, imaging and digitizing should take place in a secure, well-lit, dust-free and climate-controlled setting (minimum requirements for climate control include air-conditioning and a dehumidifier). Equipment and supply acquisitions will vary according to each country's data rescue needs and the methods selected for use. Table 4 can be used to begin planning the details of a climate data rescue project.

Choosing the appropriate personnel is critical to the success of a data rescue project. A climatologist is ideal for searching for climate data and making decisions on organizing an archive. Imaging procedures are best established by someone interested in preservation, possibly a person with an interest in photography. Individuals who are careful and pay attention to detail make the best imagers and digitizers, since speed and accuracy are important for these tasks. Several people may alternate between imaging and keying to reduce the boredom of repetitive work and to remove the single point of dependency on an individual. Both imaging and digitizing proceed more efficiently when the personnel are familiar with computer technology and when the process is done in pairs: one to image and one to check for imaging quality and two to enter or trace the same page for digitizing (double-keying). It is strongly recommended that the imaging and digitizing work be done in half a day or four-hour periods as both tasks require undivided concentration.

The number of cameras and type of copy stand required depends on whether the data are presented as bound or unbound books. When photographing one sheet at a time, two cameras are suggested, one to be used only as a backup. A copy stand is used to position both the camera and lights. To photograph data in a bound volume, however, a camera stand uses two cameras. This system enables the imager to photograph two pages at a time without damaging the binding. Appendix 6 and the I-DARE portal provide information on camera stands.

Table 4. Planning considerations for data rescue missions

	Organizing	*Imaging*	*Digitizing*
Infrastructure	Well-lit room, with climate control and security	Room with sufficient electrical sockets and lighting, as well as sufficient tables for sorting paper forms	Room with sufficient electrical sockets and lighting
Equipment	Acid-free archival boxes Double-space shelves (2-m wide, 2.5-m high shelf) Stepladder	Two or more cameras (see Appendix 6 for features and accessories) Copy stand PC Book cradle Desk Technical manuals	CDMS
Supplies	Gloves, overalls, facemasks, dust clothes	CDs, DVDs and/or two of largest external hard drives USB cable or WIFI	
Data rescue personnel	Two to clean documents and organize and develop archive and develop inventories	Two to photograph and to quality check, store and update inventory	Two to digitize and update inventory
	Training in work, health and safety management in case of contaminated records		
Personnel for follow-up	Skilled person to update inventories and to monitor room conditions	Climatologist to oversee new data and metadata ingest	Climatologist to quality control and develop products

APPENDIX 4. GUIDANCE ON MAINTAINING PAPER ARCHIVES

Paper archives should be kept in a dedicated storage room, preferably not in a basement or attic. The room should be kept clean of dust, dirt, mould and pests. Sturdy shelving and acid-free archive boxes should be used to keep documents off the floor. The room should be well-lit and boxes should be labelled and easily accessible. A dedicated person should regularly check the condition of the archival room.

Keep all paper documents at least until (a) imaging has been completed; (b) all images have been checked for quality and completeness; and (c) the images have been duplicated. It is important to make multiple electronic copies of the inventories and entire image archive and to store these in several different off-site locations.

It is preferable to always keep the paper copies:

(a) Paper copies may have indications of measurement or instrument errors (pencilled or penned slashes) that are not readable on the electronic copies;

(b) The original copies with correction marks and notes are important for verification of values;

(c) Electronic copies can be lost due to computer failure;

(d) Paper copies thus far have had a longer lifespan than digital and electronic media.

Maintaining the paper archive as a backup is prudent.

If there are no space limitations, the paper archive should remain at the NMHS. If resources do not permit continued storage of paper archives, national or university archives and libraries should be approached. A country's national archives may set legally binding standards about how long data records (both paper and electronic) should be kept. If these national institutions are not interested, international archives should be contacted.

APPENDIX 5. PRIORITIZING CONSIDERATIONS FOR IMAGING AND DIGITIZING DATA

While it might be ideal to image and key all data at once, the reality is that the archival process will take time and the process must begin somewhere. A start might be made by filling in gaps of existing time series. The International Data Rescue Meeting (Geneva, 11–13 September 2001) established that the highest priority for rescue of both historic and current climate data should be that data:

(a) Are of high quality;

(b) Are of importance nationally, regionally or globally and are at risk of loss;

(c) Would fill in gaps in established datasets.

1. IMAGING

When imaging paper documents, it seems logical to begin with the oldest, rarest or most fragile documents, one station at a time. Since paper and ink quality has declined in recent years, however, newer documents may need to be imaged sooner. When prioritizing the imaging of charts, one might begin with the charts of most importance to the NMHS (perhaps rainfall, pressure, wind or sunshine charts). All pages in a book (e.g. a monthly register) should be imaged to ensure that all metadata available (including site and data descriptions and photographs, instrument information, calibrations, observing practices and quality-control procedures) are rescued.

2. DIGITIZING

As digitizing data is time-consuming, prioritization is important. The factors below are to be considered when digitizing surface observations:

(a) To show proof of concept to potential funders, it may be advantageous to select the station with the longest continuous records or to concentrate on an unusual year that exhibits a particularly topical climate feature (such as a drought or a flood) as a way to begin;

(b) It might be prudent first to key stations that have some data already digitized that are still operational or that have the longest period of record;

(c) Fewer data are available worldwide in the pre-1950s period. In individual countries, however, any data prior to 1990 might be crucial;

(d) Location is another factor to consider, i.e. it might be beneficial to select stations that represent different climatic regions within a country or socioeconomically significant locations;

(e) Data availability might be another consideration. A station with the most reliable data and least amount of missing data might be selected first;

(f) Neighbouring stations, however – even with short duration – are important to corroborate measurements;

(g) For current or historical data, the temporal resolution of the data (monthly, daily, sub-daily) or the need for particular weather parameters could determine the priority for digitizing data;

(h) If a climatology of diurnal changes in temperature, precipitation or wind is desirable, then chart data or hourly (sub-daily) observations are crucial;

(i) To examine Essential Climate Variables, select meteorological variables that can be converted into monthly values such as temperature and precipitation (Bojinski et al., 2014);

(j) For areas where there is a spatial gap in the data, individual stations might be selected;

(k) For climatologies of thunderstorms, hail, tornadoes, duststorms, fog and ice storms, non-instrumental visual weather observations are required.

Surface observations are usually considered for imaging and digitizing, but upper-air soundings, surface and upper-air charts and marine observations are also of great importance.

(a) Individual soundings provide valuable information on changes in temperature, pressure and moisture with height;

(b) As individual sounding data may be lost, such information might only be derived from upper-air charts;

(c) Surface charts can also be quite important as original forms have sometimes been lost after surface station data were plotted.

These are the general guidelines and considerations for setting data rescue priorities. Each country or centre, however, must decide its own priorities based on the primary needs of the country and the condition of the climate data records. Once a station and year has been selected, it is wise to test the imagining and digitizing procedures. This will help determine staffing requirements and how to streamline the process.

A final reminder on digitizing forms: it is best to key all data on a given page, including metadata (since it is unlikely that there will be another opportunity to return to that page and key any omitted data). This can best be done by creating a template for digitizing the various form types. In all cases, the metadata, the description of the data and the observing practices, (observer name, observation time), the location (latitude, longitude) and site characteristics must be rescued and digitized together with the imaged or digital values.

APPENDIX 6. ELECTRONIC IMAGING TECHNIQUES

Creating an electronic image is highly desirable, because it reduces handling of the paper or microfilmed medium. It also allows the image to be resized and enhanced to make the writing more legible when key entering the data. Typically, JPEG images are suitable. Some sources suggest an image size of 1 600 x 1 200 pixels is sufficient.

1. SCANNERS

Scanners are often considered for imaging paper documents. The two general types are flatbed scanners (frequently used for books) and feeding scanners (often used for individual pages, such as charts). Scanners are an attractive alternative to cameras as they may reduce the amount of equipment required for imaging. There are a number of drawbacks to scanning, however:

(a) A flatbed scanner can damage book bindings, or may occasionally result in the data closest to the binding being out of focus or not imaged. This could be avoided with a camera set-up;

(b) Microfilm and microfiche require specially designed scanners;

(c) Scanners have many movable parts and usually require maintenance and a license;

(d) Specialized training is often required with the more complex scanning machines. Because of the training issue, changes in personnel have brought scanning to a stop in some projects. In other cases, inadequate training and licensing issues have prevented the use of scanners for imaging;

(e) Scanners are generally slow and cumbersome, except for very expensive systems.

2. CAMERAS

Cameras are used most often for creating digital images. They are effective with both bound and loose paper copies, as well as for oversized documents, such as maps. Digital cameras are commonplace and even automatic cameras can provide reliable images, especially with a constant source of good lighting. With the use of a stationary camera stand and lighting system, a remote control for taking pictures and an alternating current (AC) adapter, little needs to be done except change pages, making this process quick and efficient. In addition, inexpensive software is now available to retitle the digital images and move the image to a computer. The archival of images is also very rapid.

3. CAMERA FEATURES

Equipment needed to set up a photographic imaging workstation is summarized in Table 5. First, it is necessary to provide a table with a computer and the camera set-up (including a tripod or camera stand). An adequate number of grounded electrical sockets is crucial. Although natural lighting or sunlight are ideal for imaging – taking care that the light is diffused and shadows are avoided – a room with natural lighting is not always available. Thus, a camera set-up often includes two or four lights. A tripod or camera stand placed in conjunction with the lights allows the camera to be kept stationary so that multiple pictures can be taken quickly. It is important to test the lighting frequently, since some paper media are glossy paper which reflect light back into the camera lens, obliterating some of the imaged data. A remote-control feature allows a picture to be taken without the photographer to be positioned alongside the camera and also reduces the chance of jarring the camera body. A computer to hold the digital images is crucial.

Table 5. Basic camera infrastructure requirements

1	Table
2.	Electrical sockets
3.	Camera stand/tripod
4.	Lights
5.	Camera and accessories
6.	Computer

Cameras generally drain batteries in a relatively short time. An AC adapter allows an unlimited number of pictures to be taken and saves considerable time and expense in purchasing and replacing batteries. If an AC adapter cannot be used with the available camera, then spare, rechargeable batteries and a battery charger are recommended. Importantly, if an AC adapter is not available and batteries must be used, the camera design should allow the batteries to be replaced without removing the camera from the camera stand. Most digital cameras have plastic threads to hold the camera on the camera stand. These threads wear out after a few dozen camera removals for battery replacement.

Software and a USB cable or access to wireless communication (WIFI) is needed to automatically transmit the image from the camera to the PC. If such software allows the mouse to engage the camera shutter, a remote-control cable may not be necessary. If a USB-port or WIFI is not available, the largest memory card obtainable should be used. Digital cameras often include software to generate filenames for each image, and transmit photographs directly to a computer. Such software saves the photographer and the archivist a considerable amount of time. Depending on the camera available, a zoom lens and the ability to change camera settings is useful. Table 6 provides a summary of the desirable camera accessories that allow for an efficient process.

Consider the size of the image, particularly when using a memory card. Depending on the quality of the image, it is likely that the largest possible image is not required and that a medium-sized image may be sufficient. It has been found that even a 4-megapixel camera is enough to produce JPEG files with enough resolution to result in readable alphanumeric parameter values. Currently, even mobile phone cameras are about 12 megapixels. It must be remembered that excess fine resolution, while not needed for alphanumeric data recovery, may be needed for use with automatic computer digitizing programs. Typically, photographs are taken in colour, not black and white. Colour allows corrections and other notations found in original notebooks to be read more easily.

Table 6. Camera accessories to facilitate an efficient process

1.	A copy stand that holds a camera and 2 to 4 lights
2.	An AC adapter to power the camera
3.	A remote control for taking pictures or camera software such as in item 4
4. 5.	Camera software to take and transfer photographs directly to a computer with a USB cable or WIFI Alternatively, the largest memory card available
6.	Ability and software to rename the image files
7.	A zoom lens if readily available and macro-lens for close-ups
8.	Ability to change camera settings (via camera or computer)
9.	Glare-free glass to gently flatten bound records

Table 7. Common camera features to ensure picture quality

1.	Ability to select size of image
2.	Ability to cancel flash
3.	A macro close-up mode
4.	Ability to adjust the film speed sensitivity
5.	Ability to adjust the white balance

Other common camera features to consider are the ability to cancel the flash, a macro for close-up mode, the ability to adjust the film speed sensitivity, and the ability to adjust the white balance. Table 7 summarizes common camera features to ensure picture quality.

APPENDIX 7. OTHER DIGITIZATION METHODS

The following is a general discussion of digitization methods other than keying by individuals. All methods require quality control, since errors can occur with any transcription.

1. OPTICAL CHARACTER RECOGNITION, INTELLIGENT CHARACTER RECOGNITION AND INTELLIGENT WORD RECOGNITION

Optical Character Recognition is a commonly available software technology used to digitize typewritten text, but cannot handle handwritten observations. Intelligent Character Recognition (ICR) targets individual printed handwritten characters and allows for different fonts and styles of printing. ICR works best for reading structured forms and can achieve upwards of 97% accuracy. Intelligent Word Recognition translates whole handwritten words, either printed or cursive. Historical forms are generally not ICR/OCR-compliant, however, and will not work with current ICR/OCR technology. These latter two methods use machine-learning technology to update recognition databases automatically for new handwriting patterns.

At the time of writing, these technologies were prohibitively expensive and, as with all digitization processes, required quality control to ensure an accurate translation. These technologies are still under development and are likely to become more reliable over time. At present, digitization by keying is used most often in the belief that the human mind can best translate handwritten materials.

2. CROWDSOURCING

With crowdsourcing, information is processed by online volunteers, often referred to as "citizen scientists". A number of established and successful climate-related programmes employ crowdsourcing, notably OldWeather and the Community Collaborative Rain, Hail & Snow (CoCoRAHS) programme. Volunteers from the OldWeather project key text from ships' logs (primarily from the early years of the 20th century) into an online form. Each ship's log is keyed by three individuals and their results compared for the final product (see the I-DARE portal for a link to OldWeather). Volunteers for CoCoRAHS measure precipitation from a 10-cm (four inch) raingauge, usually near their home, and key these daily precipitation amounts into an online form. These are placed into a database and mapped.

While crowdsourcing uses volunteers for keying, there are costs. Time, talent and funding resources are required for the development and maintenance of the infrastructure for web-based data ingest, database development, quality control, developing training instructions for keyers/observers and outreach to attract potential volunteers. As with all other technologies, crowdsourcing requires quality control.

3. CHART DIGITIZATION

To digitize charts, observers usually transcribe values from the chart into a tabular form – a tedious task, especially when charts have rapid fluctuations such as anemographs or multiple lines on a single chart. While fine time resolutions can be read from a chart, peak or daily values only are often reported. With chart digitization software, however, an individual traces a line, marking points from a chart, and the points are converted to digital numbers. This allows the temporal resolution of data values from minutes to hours. In the 1990s, a number of programs to digitize precipitation (pluviograph) data were developed by individual groups, but these can only be used with certain computers and operating systems. The source code of this software is available but not maintained (e.g. updated for newer programming languages and computer technologies) and would have to be adapted by a software technician. In addition, persons maintaining and using the software need to be trained.

Recently, several new digitizing software programs have been developed and are in use. Features of these programs include the ability to compute rainfall amounts on different timescales (hours to minutes); read a variety of chart types; and handle multiple lines on a single chart. Documentation on this software can be found on the I-DARE portal. Some of these programs are available for use at no cost with the stipulation that the data be placed in a free and open database available to all.

APPENDIX 8. DIGITIZATION ERRORS AND QUALITY ASSURANCE PROCESS

The following is a brief overview of quality control of digitization errors specifically relating to keying. Some or all of these quality-control features may be available on the CDMS of the Member country. If so, as data are keyed into the CDMS, errors are flagged and later corrected. If the CDMS does not incorporate these features, it is strongly suggested that they be considered. The initial quality control of newly key-entered digital data is centred on keying, instrument, observer and metadata errors, not on the more refined analyses performed during homogenization, sometimes called "delayed mode" quality control. The keyed data should be considered (and flagged) as provisional until verified through quality-control tests.

The use of double-keying, templates, and 1-to-1 keying minimizes keying errors in individual values, but keying errors still occur. Problems due to metadata issues (largely related to station naming and identification number), scanned image quality, instrument malfunctions and non-standard observation practices remain challenges. In order to ensure the most faithful digital representation of data from meteorological forms, quality-control tests should be used to flag suspicious values of both the metadata (e.g. station name or identification number, year or month, element type) and meteorological variables. Suspect values are flagged for manual examination, with flags added indicating the type of error and the verification outcome.

Common keying errors include transposed numbers, missing or misplaced decimal points, missing digits (e.g. a zero), incorrect accumulation periods, and missing values. Other common errors include: (a) observer errors (e.g. non-conventional units); (b) instrument malfunction; (c) metadata problems (station id or naming convention inconsistency, incorrect year or month); (d) transposed or duplication of a particular element for a given month and station; and (e) unclear forms. These errors could indicate formatting issues and systematic data-quality issues.

Flags are used to identify the reason for each error. Once verification of the flagged error value is complete, the assessor selects the appropriate verification type for the data. Values may be treated as: (a) correct, matching the form, reasonable, and left unchanged; (b) an error, flagged but with no change to data; (c) an error, but with a replacement value specified; (d) non-matching, corrected to match the original form; (e) as an error and set to missing; or (f) as needing to be keyed, and then keyed. Usually, when a flagged value is ambiguous and could be correct, it is not changed.

Quality-control tests and procedures assure users that the digitized data accurately represent the observations recorded on the original documents. If tests are not run automatically by the CDMS, a quality-control inventory spreadsheet should be developed to track each test for each station. The outliers from each test need to be checked manually and any corrections applied at the conclusion of each test. After each quality-control step has been performed on a station, the quality-control accounting spreadsheet is updated.

Quality-control tests include range checks (on daily values and monthly totals and means), internal consistency checks (comparing measurements among the data types) and extreme value checks (made on individual values). Further information can be found in the WMO *Guide to Climatological Practices* (WMO-No. 100 (2016), section 3.4) and in Westcott et al. (2011) and other sources on the I-DARE portal.

APPENDIX 9. LESSONS LEARNED

1. ENCOURAGING GREATER COOPERATION BETWEEN FORECAST AND CLIMATE DIVISIONS OF AN NMHS

The Climate Division of an NMHS often oversees field operators at meteorological sites, quality-controls national climate data (hourly, daily tables) and keys in a portion of the data into its own CDMS. Typically, coded SYNOP data are sent by the weather observers using the WMO Global Telecommunication System but these digital data are often available only to the Forecast Division of the NMHS, and often only in real time. Frequently, only the handwritten hourly and daily data tables are sent to the Climate Division, not the coded digital data. Furthermore, if the Forecast Division archives the decoded weather data from the SYNOP stations or from the automated stations, they are often not available to the Climate Division. As a result, the Climate Division must obtain data for its country from the international databases. It would be ideal if the transmitted digital data were archived in a central database accessible to both the forecasters and the climatologists from which all products are drawn. Better coordination between these groups would allow the Climate Division to have a more up-to-date and, most importantly, a more complete dataset for use in developing climatological analyses. It would also allow a comparison with the printed reports for extreme value validation.

2. REVIEWING CURRENT PRACTICES OF DIGITIZATION, METADATA COLLECTION AND QUALITY CONTROL; TRAINING REQUIREMENTS

The Climate Division of an NMHS often develops its own method of data keying and quality control that evolves according to the available staffing, communication and computer resources and the number/type of meteorological stations within the country. It is advisable to regularly review current practices and incorporate relevant WMO Guidelines into the structure for inventorying, keying and quality control of current data. See Best Practice section of the I-DARE portal.

It is advisable to review any insights gained during the execution of data rescue processes that might improve the data rescue or current data ingest process or that might lead to locating other sources of data in need of being rescued. Climate data are only as good as the observers and analysts.

There will be new, unfamiliar equipment and software for the individuals who undertake data rescue work. Training and experience will be needed in the use of digital cameras, scanners, digitizing software, accessing databases and performing quality control. As people transition from job to job, multiple persons should be trained with "how to" manuals as they are learning the process to ensure that there is no single point of failure. These manuals must be revised as programs and users change.

3. DEMONSTRATING BENEFITS OF DATA RESCUE

The reorganizing, imaging, digitizing and quality-control steps of rescuing data can be expensive and time-consuming. This is probably one of the reasons why data rescue has not yet been implemented in many NMHSs. Beginning and sustaining a data rescue programme begins with the NHMS, which needs to make a clear case for the need for data rescue. This could include the identification of practical questions that can be answered by improved climate data. For instance, historical climate data would allow the NHMS to answer a minister's hydrometeorological questions accurately and quickly regarding the frequency of droughts or heavy rainfall events, or providing the direction of the strongest wind over the past 100 years so that a new US$ 10 000 000 airport runway is oriented in the right direction. Such information provides a "value-added" product that NMHS managers could charge for, bringing in additional revenue to the NMHS. In addition, organizing climate data could result in reclaiming hundreds of square metres of NMHS floor space without throwing out valuable historic records. The increased floor space could be converted to additional offices or training rooms. Finally, if the NMHS is willing to share the rescued and digitized data, it would be recognized internationally as a leader in the field.

Outside support over a number of years is often required for data rescue activities. Appendix 10 may aid in providing further justification for data rescue. Individuals and groups who have already undertaken data rescue can be advocates, who may be able to provide advice and argue for funding data rescue efforts.

Consultation with others interested in data rescue can provide technical information and possible solutions to unexpected problems along the way. The I-DARE portal will provide some technical information and possible contacts. NMHSs may contact the WMO Secretariat so that aid from other Members can be sought.

APPENDIX 10. APPLICATIONS TO SHOW VALUE OF DATA

1. VALIDATING EXTREME VALUES OF METEOROLOGICAL PARAMETERS AND PROXY DATA

It is thought that a changing and more variable climate will result in more extreme values in temperature, precipitation, pressure and wind in the future. These future values cannot be validated as “record-breaking” without knowing the extremes that have occurred in the past. Preserving original documentation and digitizing old weather records will allow easy verification of current and future extreme values. With verifiable values of weather parameters, more will be known about the overall distribution of extreme values and if the distribution of values is changing. Past climate studies in times or regions with little or no available meteorological data have been based on proxy data (tree ring, ice core). Newly rescued data are extremely valuable in verifying the validity of these methods.

2. INCORPORATE IN STUDIES, LINK TO RESEARCH PROJECTS

The best way to show the value of recovering historical climate data is by using it in research studies or having others do so. Long time series allow for the identification and examination of the severity and frequency of past events, such as droughts and floods, heatwaves and cold spells, tornadoes, tropical cyclones and hail- and duststorms. While meteorologists and climatologists are greatly interested in examining the causes of weather phenomena, others are interested in what impact these phenomena have had on society and even on the direction of a country’s history. The impacts and human reactions to these past events can aid in the planning for future extreme weather events. Using historical data to research past climate events could greatly engender enthusiasm in potential funders, especially if these same phenomena are currently happening or are expected to occur in the future. Being able to show a (changing) climatology of these features could also help predict their future probability. Complete data time series also help tune and verify climate models.

3. MAKING DATA AVAILABLE TO USERS

It is important to make the climate data available and easily accessible for users. Providing reliable historical data allows scientific researchers to engage in scientific exchange and may permit access to student assistance. Agriculture and water managers would be good allies in promoting the need for climate data rescue and they may have records or resources of their own to contribute. These managers would likely find the rescued data essential in assessing the sensitivity of natural and man-made systems to climate events and extremes. Furthermore, the development of products that would be of general interest to the public (such as the climatology of weather on national holidays) might engender interest in weather and help develop groups of “citizen scientists".

APPENDIX 11. CHECKLIST FOR CLIMATE DATA RESCUE

1. Search and locate historical datasets: paper, microfilm/microfiche and digital data in local, regional, national, international libraries and climate centres

2. Organize, clean and store paper, microfilm/microfiche and place in safe, temperature- and humidity-controlled rooms in labelled archival boxes on shelves or in filing cabinets

3. Create an electronic inventory of paper/microfilmed holdings, including station, year, month, media type, form type and box or file drawer location

4. Create an electronic image inventory spreadsheet of what has been imaged and not imaged, validated and not validated

5. Create a digital data electronic inventory spreadsheet of what has been digitized and not digitized, quality-controlled and not quality-controlled

6. Prioritize the data to be imaged and the data to be digitized

7. Review current digitization and quality-control steps to see how the new digitized data stream can fit into the NMHS's workflow, and whether current digitizing and quality-control steps need revision

8. Determine computer (CDMS), personnel, space, imaging, digitizing needs

9. Develop a workplan for imaging and digitizing

10. Obtain required equipment and hire personnel, if necessary

11. Develop an imaging process that includes imaging and archiving the data with use of inventory spreadsheets and quality control of the images and metadata

12. Image all copies of data and metadata and archive the images appropriately

13. Include image inventory in directories and on CDs

14. Duplicate image archive and store in off-site location

15. Retain original paper/microfilm copies

16. Develop a digitizing process that includes template and database development, an updated inventory, and quality control of the data and metadata

17. Review imaging and digitization processes as work progresses to determine if there are ways to improve the process

18. Cross-check inventory of imaged and digitized data with the period of record of stations to ensure completeness of process

19. On a daily basis, back up computer/CDMS with digital data

20. Update media, inventory and digital data inventories regularly with current data and metadata

21. Create multiple copies of digital data and place in multiple locations

22. Develop a technology migration plan and ensure regular migration of all electronic data (imaged and digital) to new storage media

23. Generate climatic products from quality-controlled data

24. Make data available to users in a convenient, accessible form

APPENDIX 12. GLOSSARY

ASCII. American Standard Code for Information Interchange is the most common format for text files in computers.

Autographic charts. Charts recorded mostly by a clockwork drum automatically producing meteorological measurements such as temperature, wind direction and speed and atmospheric pressure (same as stripcharts)

CDMS. Climate Data Management System

Copy stand. Structure that holds cameras and lights in position for photographing documents

DARE. Data rescue. Used for WMO activities in climate data rescue.

Digital data. Data that have been keyed into a 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 or provided by a CDMS.

Digitization. A process to transcribe analogue data into digital form for processing by a computer. Usually done by keying text data or by tracing points from a stripchart.

Electronic inventory. Spreadsheet with information on data archives

Image. A picture form of the climate data document captured either by a digital camera or scanner

Keying. Data entry or typing data into a computer via keyboard

Media. Paper, microfilm, microfiche, CDs/DVDs, tapes, hard drives or computers upon which data are written.

Metadata. 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 to enable the preservation of the intellectual content of the digital object

NMHS. National Meteorological and Hydrological Service

Preservation Keeping the data safe from harm or loss

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 or another storage media without changing format

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 developing inventories. Referred to here is the spreadsheet software either by Microsoft EXCEL or Lotus 1-2-3.

Stripcharts. Charts recorded mostly by a clockwork drum automatically producing meteorological measurements such as temperature, wind direction and speed and atmospheric pressure (same as autographic charts)

APPENDIX 13. REFERENCES

Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons and M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Amer. Meteor. Soc.*, 95:1431–1443.

Brunet, M. and P. Jones, 2011: Data rescue initiatives: bringing historical climate data into the 21st century, *Climate Research*, 47:20–40.

Westcott, N.E., K. Andsager, L. Stoecker, M.L. Spinar, R. Smith, R. Obrecht M.C. Kruk, R.T. Truesdell and D. O'Connell, 2011: Quality control of 19th century weather data. Midwestern Regional Climate Center Contract Report 11–04, Illinois State Water Survey, 68 pp.

World Meteorological Organization (WMO), 2003: *Guidelines on Metadata and Homogenization*. WMO/TD-1186, WMO, Geneva.

———, 2004: *Guidelines on Climate Data Rescue*. WMO/TD-1210, WMO, Geneva.

———, 2009: *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation*. WMO-TD 1500, WCDMP-No. 72, WMO, Geneva.

———, 2010: *Guide to Metrological Instruments and Methods of Observation.* (CIMO Guide), WMO–No. 8 (update of 2010 edition).

———, 2016: *Guide to Climatological Practices*. WMO–No. 100 (update of 2011 edition), WMO, Geneva.

For more information, please contact:

World Meteorological Organization

7 bis, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland

Communication and Public Affairs Office

Tel.: +41 (0) 22 730 83 14/15 – Fax: +41 (0) 22 730 80 27

E-mail: cpa@wmo.int

public.wmo.int

JN 161782