

CHAPTER 2. OBSERVATIONAL CLIMATE DATA AND MANAGEMENT

2.1 INTRODUCTION

For thousands of years historians have recorded information about the weather. In the past, however, this information was often based on accounts from other people and was not drawn from the historians' personal observations. Such accounts may have been vague, truncated or affected by memory lapses. This type of weather information was embedded within an immense array of other kinds of information, and much of it is contained in national libraries and archives. National archives of meteorological information are a relatively recent phenomenon, with the earliest typically being established during the first half of the twentieth century. Since the 1940s, and especially following the establishment of WMO, standardized forms and procedures have gradually become prevalent, and national meteorological-specific archives have been designated.

The quantification of climatological data developed as improved instrumentation facilitated the observation of continuous, as well as discrete, variables and the recording of appropriate values in journals or logbooks. Until the 1970s, these original forms constituted the bulk of all the holdings of climatological information at most collection centers. These centers may have been a section of the local or national government or the central office of an industry such as mining, agriculture or aviation. Gradually, the climatological data-gathering activities affecting national life were assembled within a concerted programme of observation and collection to serve national and international interests. Since the late twentieth century, most weather information has been transmitted digitally to centralized national collection centers primarily to support weather forecasting but more recently for a broader range of purposes. It has been common practice, however, to rely on the original observing documents to obtain older historical data for the creation of the climate record in climate centers around the world.

The collection, transmission, processing and storage of operational geophysical data, however, are being dramatically improved by rapid advances in computer technology, and archives are increasingly being populated with data that have never been recorded on paper. The power and ease of use of computers, the ability to record and transfer information electronically, and the development of international exchange mechanisms have given climatologists tools to rapidly improve the understanding of climate. Every effort should be made to obtain, in electronic digital form, a complete collection of all primary observed data. Collection of data electronically at the source allows rapid and automatic control measures to be applied, including error checking, prior to the data's transmission from the observation site.

The management of the vast variety of data requires a systematic approach that encompasses paper records, microform records (where relevant) and digital records. *The High-quality Global Data Management Framework for Climate* [(HQ-GDMFC), [WMO-No. 1238](#)] provides definitions, standards and recommended practices for managing data for climate purposes. It also provides the basis for dataset stewardship assessment (see 2.5). The framework and specifications for a systematic approach are given in the *Manual on the WMO Information System* ([WMO-No. 1060](#)). This publication and the manuals and guides it references are designed to ensure adequate uniformity and standardization of data, information and communications practices, procedures and specifications employed among WMO Members.

This chapter discusses general concepts and considerations for observing and managing climate data. Detailed information and guidance relating to specific topics can be found in the reference material listed in the last section of this chapter.

2.2 OBSERVATIONS

The WMO Integrated Global Observing System (WIGOS) provides a framework for the integration and sharing of observational data from National Meteorological and Hydrological Services (NMHSs) and other sources. Although aimed primarily at improving the WMO observing systems, it also interfaces with co-sponsored and non-WMO observing systems. The current goal is to facilitate the production of weather and climate services and products for the five initial priority areas of the Global Framework for Climate Services (GFCS) – agriculture and food security, disaster risk reduction, energy, health and water. WIGOS achieves interoperability and compatibility through the application of internationally accepted standards and best practices. Data compatibility is supported by the use of standardized data representation and formats. WIGOS aims to improve the quality and availability of data and metadata in order to develop capacity and to improve accessibility.

A key component of WIGOS is WMO Global Observing System. It is a coordinated system of land and space-based observing subsystems that provide in a cost-effective way high-quality, standardized meteorological and related environmental and geophysical observations. WMO Global Ocean Observing System is a similar coordinated system of marine and ocean observing networks. Data from these networks, as well as from other land, marine or satellite networks, when transmitted to central processing locations, result in the timely preparation of weather and climate analyses, forecasts, warnings, climate services, and research for all WMO programmes and relevant environmental programmes of other international organizations.

This guidance on observations specifies the elements needed to describe the climate and the stations at which these elements are measured, the required instrumentation, the siting of stations, network design and network operations. The guidance is based on *Guide to Instruments and Methods of Observation* ([WMO-No. 8](#)), *Guide to the Global Observing System* ([WMO-No. 488](#)), *Guidelines on Climate Observation Networks and Systems* ([WMO/TD-No. 1185](#)) and *Manual on the WMO Integrated Global Observing System* ([WMO-No. 1160](#)). Each edition of WMO-No. 8 has a slightly different emphasis so it is advisable to retain older editions for reference. Cross references to other WMO publications containing more detailed guidance are provided in the sections below. Guidance is also based on ten climate monitoring principles set forth in *Report of the GCOS/GOOS/GTOS Joint Data and Information Management Panel* ([WMO/TD-No. 847](#)) and in [WMO-No. 1160](#) :

- (a) The impact of new systems or changes to existing systems should be assessed prior to implementation;
- (b) A suitable period of overlap for new and old observing systems is required;
- (c) The details and history of local conditions, instruments, operating procedures, data processing algorithms, and other factors pertinent to interpreting data (metadata) should be documented and treated with the same care as the data themselves;
- (d) The quality and homogeneity of data should be regularly assessed as a part of routine operations;
- (e) Consideration of the needs for environmental and climate monitoring products and assessments should be integrated into national, regional and global observing priorities;

- (f) Operation of historically uninterrupted stations and observing systems should be maintained;
- (g) High priority for additional observations should be given to data-poor areas, poorly observed parameters, areas sensitive to change, and key measurements with inadequate temporal resolution;
- (h) Long-term requirements should be specified to network designers, operators and instrument engineers at the outset of system design and implementation;
- (i) Enabling research observing systems to perform long-term operations in a carefully planned manner should be promoted;
- (j) Data management systems that facilitate access, use and interpretation of data and products should be included as essential elements of climate monitoring systems.

These principles were established primarily for surface-based observations, but they also apply to data for all data platforms. Additional principles specifically for satellite observations are:

- (a) Constant sampling within the diurnal cycle (minimizing the effects of orbital decay and orbit drift) should be maintained;
- (b) Overlapping observations should be ensured for a period sufficient to determine inter-satellite biases;
- (c) Continuity of satellite measurements (elimination of gaps in the long-term record) through appropriate launch and orbital strategies should be ensured;
- (d) Rigorous pre-launch instrument characterization and calibration, including radiance confirmation against an international radiance scale provided by a national metrology institute, should be ensured;
- (e) On-board calibration suitable for climate system observations should be ensured and associated instrument characteristics should be monitored;
- (f) Operational generation of priority climate products should be sustained, and peer-reviewed new products should be introduced as appropriate;
- (g) Data systems needed to facilitate users' access to climate products, metadata and raw data, including key data for delayed-mode analysis, should be established and maintained;
- (h) Use of functioning baseline instruments that meet the calibration and stability requirements stated above should be maintained for as long as possible, even when these exist on decommissioned satellites;
- (i) Complementary *in situ* baseline observations for satellite measurements should be maintained through appropriate activities and cooperation;
- (j) Random errors and time-dependent biases in satellite observations and derived products should be identified.

2.2.1 Elements

A climatic element is any one of the properties of the climate system described in section 1.2.2. Measurements or observations of these elements are made as part of a variety of kinds of stations and networks. Details about these networks can be found in [WMO-No. 1160](#), the three volumes of the *Technical Regulations* (WMO-No. 49), in particular [Volume III](#), and *Guide to Agricultural Meteorological Practices* (WMO-No. 134). These documents should be kept readily available and consulted as needed.

Essential climate variables (ECV) are defined by *The Global Observing System for Climate: Implementation Needs* (GCOS–No. 200) as the physical, chemical and biogeochemical variables that critically contribute to the characterization of the earth’s climate. ECVs are:

- (a) Atmosphere surface – precipitation, pressure, surface radiation budget, surface wind speed and direction, temperature, water vapour;
- (b) Upper atmosphere - earth radiation budget, lightning, temperature, water vapour, wind speed and direction;
- (c) Atmospheric composition – aerosols, carbon dioxide, methane and other greenhouse gases, clouds, ozone, precursors for aerosols and ozone;
- (d) Hydrosphere – groundwater, lakes, river discharge;
- (e) Cryosphere – glaciers, ice sheets and ice shelves, permafrost, snow;
- (f) Biosphere - above-ground biomass, albedo, evaporation from land, fire, fraction of absorbed photosynthetically active radiation (FAPAR), land cover, land surface temperature, leaf area index, soil carbon, soil moisture;
- (g) Anthroposphere - anthropogenic greenhouse gas fluxes, anthropogenic water use;
- (h) Ocean physical - surface heat flux, sea ice, sea level, sea state, sea surface currents, sea surface salinity, sea surface stress, sea surface temperature, subsurface currents, subsurface salinity, subsurface temperature;
- (i) Ocean biogeochemical - inorganic carbon, nitrous oxide, nutrients, colour, oxygen, transient tracers;
- (j) Biological/ecosystems - marine habitats, plankton.

Datasets of observations of these variables provide the empirical evidence needed to understand and predict the evolution of climate, to guide mitigation and adaptation measures, to assess risks and enable attribution of climate events to underlying causes, and to underpin climate services.

It is important to recognize that datasets of some, if not all, of the elements are subject to many sources of error or discontinuities. Upper air radiosonde datasets, for example, are usually characterized by many discontinuities and biases resulting from instrument and operational procedural changes and incomplete metadata. Satellite observations have been available since the 1970s, and some have been assembled and reprocessed to create continuous records. Just as the radiosonde record has deficiencies, the satellite data also suffer from, among other things, limited vertical resolution, orbit drift, satellite platform changes, instrument drift, complications with calibration procedures, and the introduction of biases through modifications of processing algorithms.

Greater spatial and temporal coverage can be achieved with remote-sensing than with *in situ* observations. Remotely sensed data also supplement observations from other platforms and are especially useful when such observations are missing or corrupted. Although this is an advantage, there are problems in using remotely sensed data directly for climate applications. Most importantly, the short period of record means that remotely sensed data are often less useful than conventional instrumental data in inferring long-term climate variability and change. Also, remotely sensed data may not be directly comparable to *in situ* measurements. For example, satellite estimates of the Earth’s skin temperature are not the same as temperature measurements taken in a standard screen, and the relationship between radar measurements of reflectivity and precipitation amounts collected in rain

gauges may be quite complex. It is possible with care, however, to construct homogeneous series that combine remotely sensed and *in situ* measurements.

2.2.2 Instrumentation Concerns

This section provides guidance on some basic surface instrumentation and on the selection of instruments. There are several other WMO publications that are necessary companions to this Guide; they should be readily available and consulted as needed. A thorough survey of instruments suitable for measuring climate and other elements at land and marine stations is provided in [WMO-No. 8](#). Details of instrumentation needed for the measurement of agrometeorological elements are given in [WMO-No. 134](#)), and for hydrological purposes in the *Guide to Hydrological Practices* (WMO-No. 168, [volume I](#) and [volume II](#)).

When selecting instrumentation, including any associated data-processing and transmission systems, the ten climate monitoring principles (section 2.2) should be followed. Several concerns should be considered when complying with these principles:

- (a) Reliability - requires that an instrument functions within its design specifications at all times;
- (b) Suitability - for the operational environment at the station and to other equipment with which the instruments must operate;
- (c) Accuracy;
- (d) Simplicity of design;
- (e) Reasons for taking observations.

Ideally, all instruments should be chosen to provide the high level of accuracy and precision required for climatological purposes. It is also important that the instrument can continue to provide the required level of accuracy for a long period of time, as instrument “drift” can lead to serious inhomogeneities in a climate record; accuracy is of limited use without reliability. The [Measurement Quality Classifications for Surface Observing Stations on Land](#) (proposed to be included in [WMO-No. 8](#)) is intended to be used by network designers, managers and data users for assessment and monitoring of measurement quality at an existing station. The Measurement Quality Classifications scheme will help to better understand and quantify contributions to overall measurement uncertainty. The scheme is also applicable to support the design and optimization of a new system so that it delivers fit-for-purpose measurements and for supporting decisions on the purchase and evaluation of measurement systems.

The simpler the instrumentation, the easier it is to operate and to maintain, and the easier it is to monitor its performance. It is sometimes necessary to install redundant sensors to properly track performance and reliability over time. Complex systems can easily lead to data inhomogeneities, data loss, high maintenance cost and changing accuracy. There may be a variety of options for obtaining climate observations. When considering the options, it is important to compare costs of personnel, maintenance and replacement. Price negotiation is often possible with manufacturers on the basis of, for example, quantities purchased, among other things.

Autographic and data-logger equipment exists for the recording of many climatic elements. Observers should ensure that the equipment is operating properly and that information recorded on charts, for example, is clear and distinct. Observers should be responsible for regularly verifying and evaluating the recorded data by checking against direct-reading equipment and for making time marks at frequent,

specified intervals. The recorded data can be effectively used to fill gaps and to complete the record when direct observations are missed because of illness and other causes of absence from the observing station. [WMO-No. 8](#) gives specific guidance on the maintenance and operation of recording instruments, drums and clocks.

Data from automatic weather stations (AWSs) have been used to supplement manned stations and to increase network densities, reporting frequencies and the quantities of elements observed, especially in remote and largely unpopulated areas where human access is difficult. Some of the sensitivity and accuracy requirements of these automated stations are given [WMO-No. 8](#); others are being developed, especially for studies of climate variability. Notwithstanding the considerable potential for AWSs to provide high-frequency data, as well as additional data from remote locations, there are several significant costs associated with operating an AWS, including labour costs for maintenance and ensuring AWS reliability, labour availability, accessibility for installation and maintenance (including restoring service after outages), availability of suitable power sources, security of the site, communications infrastructure, and the necessity to meet WMO standards. These aspects must be carefully weighed against the significant benefits, such as a denser or more extensive network. Automatic weather stations can be powerful alternatives to manned observational programmes and sometimes are the only option, but they require a strong organizational commitment to manage and maintain them for the long-term.

Historically, most climatological data for the upper air have been derived from measurements made for synoptic forecasting by balloon-borne radiosondes. It is important for each NMHS to issue suitable instruction manuals to each upper-air station for the proper use of equipment and interpretation of data. It is also important to note that there are several issues concerning the quality of radiosonde measurements for climate monitoring and climate change detection purposes. Some of these issues are radiation errors, non-uniform spatial coverage and lack of inter-comparisons among types of radiosondes; metadata concerning instrumentation, data-reduction and data-processing procedures are crucial to utilizing radiosonde data in climate applications.

The climate monitoring principles should guide the development and operation of an upper-air observing system. An upper-air observing system may change over time with technological advances. Hence, a key requirement of the network is sufficient overlap of systems to maintain continuity and allow full comparison of the accuracy and precision of the old and new systems. Measurement systems should be calibrated regularly at the site. It is imperative that instrument replacement strategies take into account changes in other networks, such as the use of satellites. The Global positioning System (GPS) provides a good opportunity for sounding the upper atmosphere, especially the amount of humidity, which is related to the Zenith Tropospheric Delay of a GPS Signal. A world network of Terrestrial GPS Station receivers is available and well maintained. Thus, their measurement and data can be used to monitor the climate system.

The most common surface-based active remote-sensing technique is weather radar. Atmospheric and environmental conditions can adversely affect radar data, and caution should be exercised when interpreting the information. Some of these effects include returns from mountains, buildings and other non-meteorological targets; attenuation of the radar signal when weather echoes are viewed through intervening areas of intense precipitation; temperature inversions in the lower layers of the atmosphere, which bend the radar beam in such a way that ground clutter is observed where normally not expected; and the bright band, which is a layer of enhanced reflectivity caused by the melting of ice

particles as they fall through the freezing level in the atmosphere, which can result in overestimation of rainfall. Use of radar data in climate studies has been limited by access and processing capabilities, uncertainties in calibration and calibration changes, and the complex relationship between reflectivity and precipitation.

Wind profilers use radar to construct vertical profiles of horizontal wind speed and direction from near the surface to the tropopause. Although they work best in clear air, wind profilers are capable of operating in the presence of clouds and moderate precipitation. When equipped with a radio-acoustic sounding system, profilers can also measure and construct vertical temperature profiles. The climatic characteristics and geographical constraints of the area should dictate the types and/or nature of the instruments that are installed

The accuracy and efficiency of a lightning-detector network drops progressively on its outer boundaries. The detection wave will propagate without too much attenuation over distance, depending on the frequency band used, but if a lightning flash is too far away from the network (this distance varies with the stroke amplitude and the network configuration), the stroke may no longer be detected. Also, measurements of the same element should be taken simultaneously by more than one instrument for a period long enough to calibrate the instruments and increase the quality of the data.

Many long-distance aircraft are fitted with automatic recording systems that report temperature and wind, and in some cases humidity, regularly while *en route*. Some aircraft record and report frequent observations during take-off and descent to significantly augment the standard radiosonde data, at least throughout the troposphere. Such data are assimilated into operational meteorological analysis systems and, through programs of reanalysis, ultimately contribute substantially to the broader climate record. Aircraft measurement calculation uncertainty should be evaluated along with the effects of sampling interval and averaging time. Examination of typical time series of vertical acceleration data often indicates a high variability of statistical properties over short distances. Variation of air speed for a single aircraft and between different aircraft types alters the sampling distances and changes the wavelengths filtered. While not as precise and accurate as most ground observing systems, aircraft data can provide useful supplemental information to meteorological databases.

The Space Programme coordinates satellite data processing and dissemination activities based on common standards. A focus of the programme is the integration of space weather observations, through review of space- and surface-based observation requirements, harmonization of sensor specifications, and monitoring plans for space weather observation. Some of the objectives of the programme are enhancing the accessibility of current and next-generation satellite data and products and responding to user needs, promoting data exchange through common standards and the WMO Information system (WIS), and stimulating coordinated data processing with traceable quality. In order to achieve these objectives, the programme has analyzed how the typical cycle of satellite system development relates to user-readiness projects. The recommended best practices are discussed in *Guidelines on Best Practices for Achieving User Readiness for New Meteorological Satellites* ([WMO-No. 1187](#)). The technology of remote-sensing is evolving rapidly. The latest documents should be referred to in processing, archiving and using remote-sensing data.

2.3 CLIMATOLOGICAL NETWORKS

A network of stations is a collection of multiple stations of the same type (such as a set of precipitation stations, radiation measuring stations or climatological stations), which are administered as a group. Each network should be optimized to provide the data and perform as required at an acceptable cost. Most optimizing methods rely on data from a pre-existing network, available over a long enough period to correctly document the properties of the meteorological fields. They are based on both temporal and spatial statistical analyses of time series. It is difficult to assess *a priori* how long the data series must be because the number of years necessary to capture variability and change characteristics may vary with the climatic element. It has been common practice to assume that at least ten years of daily observations are necessary to produce the relevant base statistical parameters for most elements, and at least thirty years for precipitation. Observed global and regional climatic trends and variability in many areas of the globe over the past century suggest, however, that such short periods of record may not be representative of similar periods to follow.

The identification of redundant stations allows network managers to explore options for optimizing the network, for example, by using the resources to establish stations at locations where observations are needed for a more effective realization of the network objectives. Network managers should take advantage of the relatively high spatial coherence that exists for some meteorological fields, such as temperature. Techniques used to evaluate the level of redundancy of information include the use of the spatial variance–covariance matrix of the available stations, multiple linear regression, canonical analysis and observation system simulation experiments. Care should be taken, however, when considering the removal of stations since redundancy and spatial coherence are important tools for assessing data quality and homogeneity.

The density and distribution of climatological stations to be established in a land network within a given area depend on the meteorological elements to be observed, the topography and land use in the area, and the requirements for information about the specific climatic elements concerned. The network should give a satisfactory representation of the climate characteristics of all types of terrain in the territory of the Member concerned (for example, plains, mountainous regions, plateaus, coasts and islands). For data used in sectoral applications within an area, there may be a need for a greater density of stations where human activities or health are sensitive to climate, and a lesser density in locations with fewer people. RBSNs of both surface and upper-air stations, and RBCNs of climatological stations shall be established to meet the requirements laid down by the regional associations.

WMO is in the process of implementing a radical overhaul of the international exchange of observational data, which underpins all weather, climate and water services and products. In some parts of the world observations are either not made or not exchanged internationally, and in other parts they are not made or exchanged frequently enough. The [Global Basic Observing Network](#) initiative will be a fundamental element of WIGOS. It will concern how the basic surface-based observing network is designed, defined and monitored at the global level. Once implemented, the availability of the most essential surface-based data will be improved.

Each Member should establish and maintain at least one reference climatological station for determining climate trends. Such stations need to provide more than 30 years of homogeneous records and should be situated where anthropogenic environmental changes have been and are expected to remain at a minimum. Information on agrometeorological and hydrometeorological networks and sites

can be found in [WMO-No. 134](#) and WMO-No. 168, [volume I](#) and [volume II](#), respectively, and additional guidance is provided in [WMO-No. 1160](#).

A country's environmental information activities are often conducted by many parties whose contributions are complementary and at times overlapping. A country benefits from environmental information collected and disseminated by both governmental agencies and non-governmental entities (including private companies, utilities and universities). Formal partnerships between the NMHS and these other parties are highly desirable for optimizing resources. Because data and information obtained from non-NMHS sources are not usually under the control of the NMHS, metadata are critical for the most effective use of the information. As for stations maintained by the NMHS, metadata on instrumentation, siting, processing procedures, methodologies and anything else that would enhance the use of the information should be obtained and documented. The metadata should also be maintained and be accessible. To promote the open and unrestricted exchange of environmental information, including weather observations, it is highly desirable that the NMHS be granted full use of all the climate data and information obtained from partnerships, without restriction, as if they were its own data. An appropriate contract or memorandum of understanding between the NMHS and other organizations may need to be drafted and signed at the senior management level.

In addition to data from standard and private networks of climatological stations, there are sometimes observational data from networks of temporary stations established in connection with research and study programmes, as well as measurements made in mobile transects and profiles. The NMHS should endeavor to obtain these data and associated metadata. Although the data may not be ideal for typical archiving, they will often prove to be quite valuable as supplementary information, for example, for investigations of specific extreme events. When these observations are collected from data-poor areas, they are highly valuable. It is also worth noting the initiatives of WMO and GCOS to maintain and protect the national long-term meteorological observations with WMO recognition of [centennial observing stations](#).

2.3.1 Siting considerations

The precise exposure requirements for specific instruments used at climatological stations, aimed at optimizing the accuracy of the instrumental measurements, are discussed in [WMO-No. 8](#). The representativeness and homogeneity of climatological records are closely related to the location of the observing site. A station sited on or near a steep slope, ridge, cliff, hollow, building, wall or other obstruction is likely to provide data that are more representative of the site alone and not of a wider area. These stations, such as lighthouses, may be of value if they have been in a stable environment for a long period of time. A station that is or will be affected by the growth of vegetation, including even limited tree growth near the sensor, growth of tall crops or woodland nearby, erection of buildings on adjacent land, or increases (or decreases) in road or air traffic (including those due to changes in the use of runways or taxiways) will provide neither broadly representative nor homogeneous data.

A climatological observing station should be sited at a location that permits the correct exposure of the instrumentation and allows for the widest possible view of the sky and surrounding country, if visual data are required. Stations should be sited on a level piece of ground covered with short grass or other bare ground conditions natural to the site; the site should be well away from trees, buildings, walls and steep slopes and should not be in a hollow. A plot size of about 9 meters by 6 meters is sufficient for outdoor temperature and humidity-sensing instruments, and an area of 2 meters by 2 meters of bare

ground within the plot is ideal for observations of the state of the ground and soil temperature measurements. A slightly larger plot (10 meters by 7 meters) is preferable if the site is to enclose a rain gauge in addition to the other sensors.

A rule used by many NMHSs is that the distance of any obstruction, including fencing, from the rain gauge must be more than twice, and preferably four times, the height of the object above the gauge. In general terms, anemometers should be placed at a distance from any obstruction of at least 10, and preferably 20, times the height of the obstruction. The different exposure requirements of various instruments may give rise to a split site, where some elements are observed from one point while others are observed nearby, with data from all the elements combined under the one site identifier.

It is critical to include notes, drawings and/or photos that describe the station exposure, and that these are updated regularly, even if just to say station exposure is unchanged. If the site's environment changes seasonally (for example, through changes in nearby vegetation) this information should also be recorded. This information is an important part of a station's metadata, and it forms a critical component of historic investigations and homogeneity assessments.

Prevention of unauthorized entry is a very important consideration and may require enclosure by a fence. It is important that such security measures do not themselves compromise the site exposure. Automatic stations will normally need a high level of security to protect against animal and unauthorized human entry; they also require the availability of suitable and robust power supplies, and may possibly need additional protection against floods, leaf debris and blowing sand.

Stations should be located at sites that are subject to such administrative conditions that will allow the continued operation of the station, with the exposure remaining unchanged, for a decade or more. For stations used or established to determine long-term climate change, such as reference climatological stations and other baseline stations in the GCOS network, constancy of exposure and operation is required over many decades.

Observing sites and instruments should be properly maintained so that the quality of observations does not deteriorate significantly between station inspections. Routine, preventive maintenance schedules include regular "housekeeping" at observing sites (for example, grass cutting and cleaning of exposed instrument surfaces, including thermometer screens) and manufacturers' recommended checks on instruments. Routine checks carried out at the station or at a central point should be designed to detect equipment faults at the earliest possible stage. Depending on the nature of the fault and the type of station, the equipment should be replaced or repaired according to agreed priorities and time intervals. It is especially important that a log be kept of instrument faults and remedial action taken where data are used for climatological purposes. This log will be the principal basis for the site's metadata and hence will become an integral part of the climate record. Detailed information on site maintenance can be found in [WMO-No. 8](#).

The nature of urban environments makes it impossible to conform to the standard guidance for site selection and exposure of instrumentation required for establishing a homogeneous record that can be used to describe the larger-scale climate. Nonetheless, urban sites do have value in their own right for monitoring real changes in local climate that might be significant for a wide range of applications. Guidelines for the selection of urban sites, installation of equipment and interpretation of observations are given in *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites*

(WMO/TD-No. 1250). Fundamental to the guidance is the need to clearly understand the purpose of making the observations and to obtain measurements that are representative of the urban environment. In many urban situations, it will be possible to conform to standard practices, but flexibility in siting urban stations and in instrumentation may be necessary. These characteristics further heighten the importance of maintaining metadata that accurately describe the setting of the station and instrumentation.

2.3.2 Scheduling considerations at manned stations

Observations at climatological and precipitation stations should be made at least once (and preferably twice) each day at fixed hours that remain unchanged throughout the year. From a practical viewpoint times of observation should fit the observer's working day, usually one morning observation and one afternoon or evening observation. If daylight saving time is used for part of the year, the observations should continue to be made according to the fixed local time; the dates when daylight saving time commences and ends must be recorded. It is desired that the times of observation should coincide with either the main or intermediate standard times for synoptic observations (0000, 0300, 0600 Coordinated Universal Time [UTC], and so on) although existing observation times should not be altered solely for this purpose. If conditions dictate that only one observation a day is possible, this observation should be taken between 0700 and 0900 local standard time.

In selecting the schedule for climatological observations, times at or near the normal occurrence of daily minimum and maximum temperatures should be avoided. Precipitation amounts and maximum temperatures noted at an early morning observation should be credited to the previous calendar day, while maximum temperatures recorded at an afternoon or evening observation should be credited to the day on which they are observed.

Times of observation often vary among networks. Summary observations such as temperature extremes or total precipitation made for one 24-hour period (such as from 0800 on one day to 0800 on the next day) are not equivalent to those made for a different 24-hour period (such as from 0000 to 2400).

If changes are made to the times of observations across a network, simultaneous observations should be carried out at a basic network of representative stations for a period covering the major climatic seasons in the area at the old and new times of observation. These simultaneous observations should be evaluated to determine if any biases result from the changed observation times. The station identifiers for the old and new times of observations must be unique for reporting and archiving. If possible, network-wide changes of this type should be avoided.

The observer should note in the station logbook and on the report forms the nature and times of occurrence of any damage to or failure of instruments, maintenance activities, and any change in equipment or exposure of the station, since such events might significantly affect the observed data and thus the climatological record. Where appropriate, instructions should be provided for transmitting observations electronically. If mail is the method of transmission, instructions for mailing should be provided to the station, as well as pre-addressed, stamped envelopes for sending the report forms to the central climate office.

2.3.3 Observer competency

Guide to Competency ([WMO No. 1205](#)) provides the general framework for competencies, and *Compendium of WMO Competency Frameworks* ([WMO- No. 1209](#)) describes the framework specifically for personnel performing meteorological observations. It describes the knowledge and skills required to perform many observing functions. Observers should be trained and certified by an appropriate meteorological service to establish their competence to make observations, maintain equipment, and perform their duties to the required standards. They should have the ability to interpret instructions for the use of instrumental and manual techniques that apply to their own particular observing systems. Guidance on the instrumental training requirements for observers is given in [WMO-No. 8](#).

Often, observers are either volunteers or part-time employees, or take observations as part of their other duties. They may have little or no training in climatology or in taking scientific observations, and thus will depend on a good set of instructions. Instructional booklets for ordinary climatological and precipitation station observers should be carefully prepared and made available to observers at all stations. The instructions should be unambiguous and should simply outline the tasks involved; they should provide only the information that the observer actually needs in order to perform the tasks satisfactorily. Illustrations, graphs and examples could be used to stimulate the interest of the observer and facilitate the understanding of the tasks to be undertaken every day. Sample copies of correctly completed pages of a logbook or journal and of a report form should be included in the instruction material available to an observer. Ideally, a climate center representative should visit the site, install the station and instruct the observer.

An observer must gain familiarity with the instruments, and should be aware in particular of the sources of possible error in reading them. The instructions should include a descriptive text with simple illustrations showing the functioning of each instrument. Detailed instructions regarding methods to be used for day-to-day care, simple instrument maintenance and calibration checks should be given. If correction or calibration tables are necessary for particular observing and recording tasks, the observer should be made thoroughly familiar with their use. Instructions should also cover the operation of computer terminals used for data entry and transmission.

Instructions must cover visual as well as instrumental observations. Visual observations are particularly prone to subjective error and their accuracy depends on the skill and experience acquired by the observer. Since it is very difficult to check the accuracy or validity of an individual visual observation, as much guidance as possible should be given so that correct observations can be made.

To complement the instruction material, personnel responsible for station management in the climatological service should contact observing stations regarding any recurring observing errors or misinterpretation of instructions. Regular inspection visits provide the opportunity to address siting or instrument problems and to further the training of the observer.

Some climate centers arrange special training courses for groups of volunteer observers. Such courses are especially useful in creating a uniform high standard of observations, as a result of the training given and the availability of time to address a wider range of problems than may be raised by a single observer at an on-site visit.

2.3.4 Quality management of climatological station operations

Guidance for on-site quality control of observations is given in [WMO-No. 488](#). Checks should be made for gross errors, against existing extremes, for internal consistency in a sequence of observations, for consistency in the sequence of dates and times of observation, for consistency with other elements and calculations, and for the accuracy of copies and of encoded reports. If there are errors, remedial action such as correcting the original data and the report should be taken before transmission. Errors detected after transmission should also be corrected and the corrected report should be retransmitted. Checks should also be made, and any necessary amendments recorded and corrections transmitted, if a query about data quality is received from an outside source. Records of an original observation containing an error should include a notation or flag indicating that the original value is erroneous or suspect. On-site quality control must also include maintenance of the standard exposure of the sensors, of the site, and of the proper procedures for reading the instrumentation and checking autographic charts. Any patterns of measurement error should be analyzed, for example, to see if they relate to instrument drift or malfunction, and summaries of data or report deficiencies should be prepared monthly or annually.

Climatological stations should be inspected to ensure the maintenance and correct functioning of the instruments and thus a high standard of observations. Automated stations should be inspected at least every six months. Special arrangements for the inspection of ship-based instruments are described in [WMO-No. 8](#).

Before each inspection, the inspector should determine to the fullest extent possible the quality of information and data received from each station on the itinerary. At each inspection, it should be confirmed that:

- (a) The observer's training is up to date;
- (b) The observer remains competent;
- (c) The siting and exposure of each instrument are known, recorded and still the best obtainable;
- (d) The instruments are of an approved pattern, in good order and verified against relevant standards;
- (e) There is uniformity in the methods of observation and procedures for calculating derived quantities from the observations;
- (f) The station logbook is well maintained;
- (g) The required report forms are sent punctually and regularly to the climate center.

Inspection reports should include sketches, photographs or diagrams of the actual observing site, indicating physical objects that might influence the observed values of the climatic elements. The reports must also list any changes in instruments and any differences in readings between instruments and travelling standards, changes in exposure and site characteristics from the previous visit, and dates of appropriate comparisons and changes. Inspectors must also be prepared to advise observers on any problems arising in the transmission of data, including automated data-entry and transmission systems.

It is of paramount importance, for determining the spatial and temporal variations of climate, that the relative accuracy of individual sensors in use in a network at one time be measured and periodically checked and, similarly, that the performance of replacement sensors and systems can be related to that of the instruments replaced. Details on calibration techniques can be found in [WMO-No. 8](#). For

climatology, it is not generally sufficient to rely upon manufacturers' calibrations and it is wrong to assume that a calibration will not drift or otherwise change with time. Records of instrument changes and calibration drifts must be kept and made available as metadata, as they are essential to the assessment of true climate variations.

During inspections of remotely sited AWSs, observations should be taken using travelling standards for later comparison with the recorded AWS output received at the data reception point. Some NMHSs have automated fault or instrumental drift detection procedures in place, which compare individual measurements with those from a network and with values analyzed from numerically fitted fields. These automated procedures are useful for detecting not only drift, but also anomalous step changes.

2.4 CLIMATE DATA PROCESSING

The basic goal of climate data processing is to capture, preserve and provide access to climate data and products for use by planners, decision-makers and researchers. Permanent archiving is an important objective. The data management system of a climate archive must provide the information to describe the climate of the domain of interest for which the archive has been established, be it national, regional or global. Many of the ultimate uses of climate data cannot be foreseen when the data acquisition programmes are being planned, and frequently new applications emerge, long after the information is acquired. The initial utilization of meteorological and related data is often only the first of many applications. Subsequent analysis of the data for many and diverse purposes leads to a significant and ongoing enhancement of the return on the original investment in the data acquisition programmes. The global climate change issue, for example, is stretching the requirements for climate data and data management systems far beyond those originally conceived when the original networks were established. To meet these expanding needs, it is critically important that climate information, both current and historical, be managed in a systematic and comprehensive manner. Conventional meteorological data are now augmented by data from a wide array of new instruments and systems, including satellites, radar systems and other remote-sensing devices, and model data, thus making effective and comprehensive Climate Data Management Systems (CDMSs) essential for modern climate centers.

[WMO-No. 1238](#) provides guidance and requirements on the development, provision, exchange and maintenance of high-quality climate datasets. The standards and recommended practices it describes are intended to ensure that the data made available for climate assessment, monitoring, applications and related services meet sustainably a minimum set of requirements with regard to quality, governance, accessibility and usability. It enables the effective development and exchange of high-quality climate data, based on a reliable, integrated, underpinning data infrastructure at the global, regional and national levels. HQ-GDMFC provides a robust data foundation for the generation of climate products and the delivery of climate services through the Climate Services Information System (CSIS) of GFCS.

Collaboration within GFCS applies to several disciplines, such as marine and hydrological sciences, for which many of the data-related activities are considered as falling under the definition of climate data. The HQ-GDMFC data scope encompasses all of the ECVs (see 2.2.1) under the WMO auspice. This includes observational data as well as data derived from climate analysis, reanalysis, prediction and projection. The procedures provided are also applicable to externally sourced data and data on socioeconomic impacts outside WMO auspices.

2.4.1 Climate database management

The primary goals of database management are to maintain the integrity of the database at all times, and to ensure that the database contains all the data and metadata needed to meet the requirements for which it was established, both now and into the future. Climate Database Management Systems (CDMSs) allow efficient storage, access, conversion and updating of many types of data, including digital and non-digital data, and by enhancing data security.

It is essential that both the development of climate databases and the implementation of data management practices take into account the needs and capabilities of existing and future data users. While this requirement may seem intuitive, information that is important for a useful application is sometimes omitted, or data centers commit insufficient resources to checking the quality of data for which users explicitly or implicitly demand high quality. In all new developments, data managers should attempt to have at least one key data user as part of the project team or to undertake some regular consultative process with user stakeholders to keep abreast of both changes in needs and any issues that user communities may have. Examples of stakeholder communities are those involved in climate prediction, climate change, agriculture, public health, disaster and emergency management, energy, natural resource management, urban planning, finance and insurance.

A CDMS should be monitored routinely to determine how well the processes that use and support the database are performing. Examples of the processes that support the data are metadata maintenance, database ingestion, quality control actions that modify the database, and information retrieval. Each process should be monitored, evaluated and, if necessary, improved. It is strongly recommended that data managers think in terms of end-to-end data management, with information on systemic data quality issues, loss of data, or other practices that harm the climate record being referred back to observation managers for rectification.

A security policy to prevent loss of or damage to the CDMS should insure that:

- (a) All personnel must be aware of their professional responsibilities;
- (b) The archives and database environment must be secured and protected against physical hazards to the records, such as fire and excess humidity;
- (c) For digital data, user-level security should be enforced with respect to the database and its components. Only a small and registered group of people should have the right to perform data manipulations such as insertions, updates or deletions;
- (d) Personnel with write access to a database must agree not to perform any transactions besides the operations and practices approved by the data manager;
- (e) All changes to data tables should have an audit trail, and controls on access to this trail should be in place;
- (f) Password security principles should be applied, including not sharing passwords, not writing passwords on paper, changing passwords regularly, and using “strong” passwords consisting of seemingly unrelated letters, numbers and characters;
- (g) All unnecessary services should be disabled on the database computer;
- (h) The database must be protected against attacks from viruses and hackers;
- (i) Regular backups must be made, noting that work done after the most recent backup will likely be lost and need to be repeated should a computer failure occur. Typically, an incremental backup should be made daily and a full backup weekly;

- (j) Every so often, typically monthly, a complete backup of the data tables should be put in a safe, secure, fireproof location, remote from the physical location of the climate database. It is common to have three copies of the same archive in different secure places and if possible in different towns or cities;
- (k) Backups of the CDMS must be performed prior to any changes to the system software, to the system design or to the applications contained in the CDMS.

An important function of a database manager is to estimate data storage requirements, including the estimation of future growth. With the increasing use of automatic stations, there is also the question of whether or not to save data at time steps smaller than the hourly time step, such as minute data. It is obvious that the ability to manage minute data or even grid point data will have important consequences on the resources needed to process these data. Account must be taken of the additional information to be included in data records (for example, data quality flags, original messages, and date and time of record updates), metadata needs and any redundancy necessary to ensure that databases can be restored. Some data types, such as those from remote-sensing, marine sensors and AWS with high temporal resolution, require large amounts of storage. Unconventional data (such as soil moisture, phenological observations and vegetation indices) may have storage needs that are different from the more traditional observations. AWS will often generate data that are relevant to the quality of observations, but are not strictly climate data (for example, information on the battery-level voltage for an AWS). The quality control process often generates values and information that may be different from the original data, so there is a storage requirement for keeping both the original data and any different data generated by quality control processes. Non-digital records should be stored in a controlled environment that minimizes their deterioration from temperature and humidity extremes, insects, pests, fire, flood, accidents or deliberate destruction.

With an ever-increasing amount of information being generated and retained, the problem arises as to whether or not to continue to store all the records in their original manuscript form. All too often, climatological records are stored in basements, sheds and other undesirable facilities. They are frequently not cataloged, and they may be inaccessible and subject to deterioration. As a means of reducing paper costs, making better use of space, and providing security for original documents, it is recommended that the manuscript data be scanned into a digital file and carefully preserved. It is important to remember that no storage medium is permanent, and therefore regular review of archival arrangements should be undertaken.

2.4.2 CDMS design

All CDMSs are based on some underlying model of the data. This model design is very important for the quality of the resulting system. An inappropriate model will tend to make the system harder to develop and maintain. In general, a database designed for current meteorological data will allow rapid retrieval of recent data from a large number of stations. By contrast, many climate data applications involve the retrieval of data for one or a few stations over a long period. It is essential to document the overall design and underlying data model of the CDMS to facilitate subsequent extension or modification as needs or technologies change. Similar considerations apply to a metadata model.

In general a CDMS has five components. The governance component refers to a consistent set of policies and governance processes needed to build a solid foundation for the establishment and management of authoritative sources of climate data and related services. The data management component addresses

the functionality required to effectively manage climate data and includes data ingest and extraction, data rescue, quality control of observations, quality assessment and management of climate metadata. The data delivery component refers to the functionality required to deliver climate data and includes data discovery (both climate data and climate metadata) and data delivery. The data analysis component involves a wide variety of analytical techniques that are applied to climate data and may result in the generation of a range of derived data products. Some examples are spatial and image analysis, homogenization and numerical modelling processes. The data presentation component represents a diverse set of techniques used to communicate climate-related information such as written reports, graphical static and interactive presentations and multimedia presentations. All of these components require the computer infrastructure functionalities to support a CDMS. Not all components are required; *Climate Data Management System Specifications* (WMO-No. 1131) gives details about CDMSs and discusses whether a functionality is considered as required, recommended or optional.

2.4.3 Data acquisition considerations

Manually observed data should be transferred to the CDMS as soon as possible by whatever means are most practical. It is advantageous to collect data at least daily because data quality is likely to be improved, the manual effort for quality control will likely decrease, technical errors will be detected faster, and there will be greater opportunities for improved access to more data. Nevertheless, the submission of data for a month is an acceptable alternative when daily data transmission is not practicable, such as when observations are recorded on paper and sent to the NMHS by mail.

Non-digital records are generally digitized during an entry process. A fundamental goal of a data-entry process is to duplicate, with a minimum of error, the raw data as they were recorded in the capture process. A key-based entry system should be efficient and easy for a data-entry operator to use. The system could also be designed to validate the data as they are entered and to detect likely errors. It is also possible to set default values for some elements, thus saving unnecessary keystrokes.

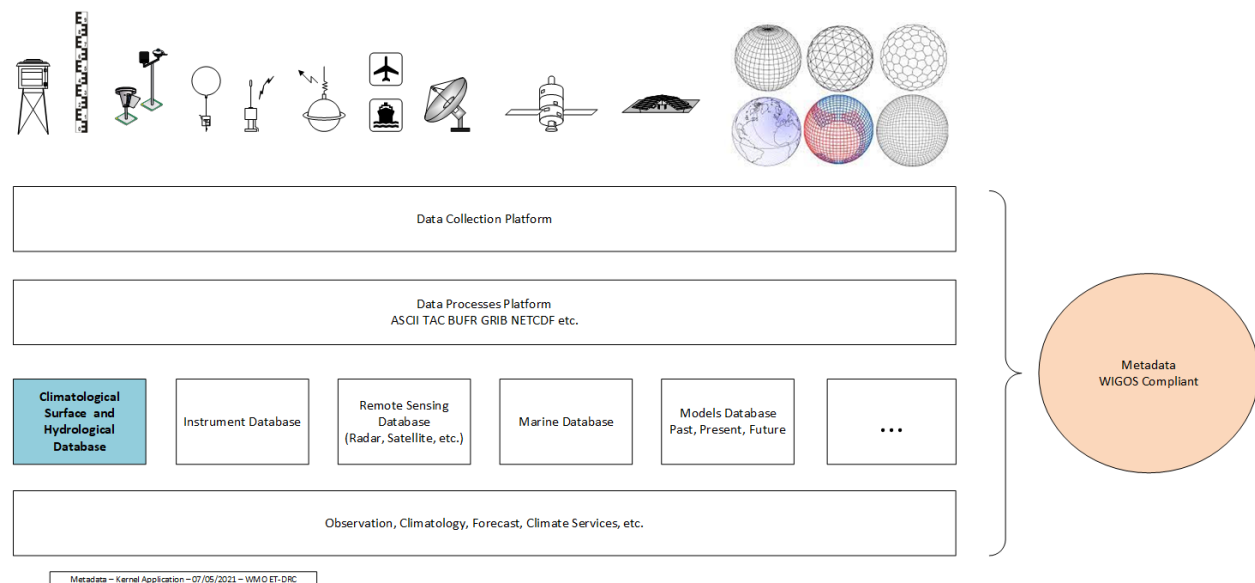
Many weather observations are recorded by institutions or organizations other than NMHSs, and acquisition of the data in their original form may be difficult. In these cases efforts should be made to collect copies of the original report forms. If it is impossible to secure either the original or a copy of the record, a note to this effect should be made in the inventory of the center's holdings, stating information pertaining to the existence and location of the data, volume available, period of record covered, stations in the network as applicable and elements observed.

It is recommended that the CDMS also contain information about media reports, pictures and other similar information beyond the traditional data and metadata. Such information could be captured by imaging the report from the print media with a digital camera or scanner; defining the date, area and type of event (such as flood, drought or heavy precipitation); identifying the media; and writing additional comments about the event.

It is important to retain both the data value that was originally received, as well as the latest quality-controlled value. The original value will likely pass through an initial automated quality control process at ingestion and, as necessary, a more extensive quality control process; even if rejected by either of these processes, it must still be retained. Some CDMSs retain not only the original and latest values, but also all modifications performed on the data and the answers to the following questions: when, why, how and who?

Another aspect of data acquisition is the recording of occurrences when data were expected but not received. Loss of data can occur as a result of situations such as inoperable instrumentation, data transmission errors and acquisition processing errors. Lost data can be reconstructed with varying levels of certainty. For example, a missing precipitation measurement can be considered to be zero when it is known from other data that local and synoptic conditions precluded the occurrence of precipitation. In other cases, lost data can be estimated with reasonable certainty (see 3.5.8). In all cases, dataset documentation should flag the reconstructed or estimated data appropriately.

The following schema shows a data flow as used frequently by NMHSs having sufficient Information Technology capacities:



2.4.4 Metadata

An adequate set of metadata must be available to inform future users about the nature of the data in the system, how the various datasets were collected, and any inherent problems. A user should be able to use the metadata to identify the conditions under which the observation (or measurement) was made, and any aspects that may affect its use or understanding. It is recommended that database management include all information that can affect the homogeneity of a dataset or series. Descriptions of and requirements for metadata are given in *WIGOS Metadata Standard* (WMO-No. 1192). Other helpful metadata information can be found in *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186). These standards are part of the WMO Integrated Global Observing contributions to co-sponsored observing systems in support of all WMO Programmes and activities (see WMO-No. 1160). Most modern metadata are available in digital form but older metadata may exist only on paper.

Metadata are needed not only for climate data but also for the CDMS itself. Each process within the system (for example, key entry or quality control) should be completely described. A history of any changes made to any part of the system (for example, software, hardware or manual procedures) should be documented and maintained. Since observing practices, quality control techniques and data-handling procedures change over time, these metadata are critical in the climatological analysis of

historical data. The analyst uses the metadata to identify and understand how a data value was observed and processed in order to separate meteorological from possible non-meteorological influences in the data record.

Inventories of the data contained in the CDMS should be prepared routinely. Stratification could be, for example, by data element, station location, or time. Lists of contents should be established and maintained to describe and define the data content of the individual files and to provide information on the codes and observational practices used. Knowing what is contained in the CDMS is important for efficient retrieval of information from the system.

2.4.5 Quality Control

The objective of quality control is to verify whether a reported data value is representative of what was intended to be measured and has not been contaminated by unrelated factors. It is important, therefore, to be clear from the outset what the readings of a particular data series are meant to represent. Data should be considered as satisfactory for permanent archiving only after they have been subjected to adequate quality control.

Procedures for quality control help assure defined data quality levels throughout the life cycle of the data and should be an integral part of an entity's quality management system. It should be noted that the procedures serve in particular to assure data quality levels that support climate applications and services. It should also be noted that quality control procedures need to be adapted to the specific climate conditions of a country and fine-tuned to fit existing and planned observational and information technology infrastructure as well as available human resources. Applied quality control procedures should be well documented and made available to data users.

If manuscript records constitute the source document, trained personnel should, upon receipt at the archiving center, scrutinize them before any digitization takes place. The forms should be reviewed to ensure proper identification (for example, station name, identifying number and location), legibility and proper recording of data (for example, to the correct precision and in the proper columns). If any problems are discovered, the observation sites should be contacted to clarify the issues or correct the problems. If resources do not permit the quality control of all data, priority should be given to the most important climate elements.

When observed data are available in digital form, the archiving center should subject them to full, elaborate quality control procedures on a regular, systematic basis. Computer software can examine all the available data and list those that fail pre-defined tests, but are not so adept at identifying the underlying problem. A skilled human analyst can often make judgments about the cause of errors and determine any corrections that should be applied, but is generally overwhelmed by the vast quantity of observations. The best technique is a combination of the two, with computer-generated lists of potential errors presented to the human analyst for further action. For datasets with very large data volumes (such as AWS data with 1-minute resolution), human intervention is likely to be impractical and quality control will be entirely or almost entirely automated.

Statistical techniques (described in Chapter 3) are invaluable for detecting errors, and in some cases for suggesting what the "correct" value should be. Objective, automated screening of data is essential when validating large quantities of data. A manual review of the automated output is needed, however, to

ensure that the automated procedures are indeed performing as expected. Graphical and map displays of data and data summaries are excellent tools for visual examinations. These techniques integrate and assimilate large quantities of data and enable a trained analyst to recognize patterns for assessing physical reasonableness, identifying outliers, noticing suspect data and evaluating the performance of automated procedures. Additional advice is in *Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications* (WMO-No 1269).

In a database a given value is generally available at different stages of quality control. The original data as received in the database must be kept, but validation processes often lead to modifications of the data. These different stages of the value are reflected in quality flags. A multitude of flags could be constructed, but the number of flags should be kept to the minimum needed to describe the quality assessment and reliability of the raw data or estimated values. A quality flag code using two digits, one for the type of data and one for the stage of validation, meets most requirements. When data are acquired from multiple sources, a third flag for the source, or provenance, of the data is often useful. Examples of “type of data”, “validation stage” and “acquisition method” codes are given in Tables 3.1, 3.2 and 3.3.

Table 3.1. Example of type of data codes

<i>Type of data code</i>	<i>Meaning</i>
0	Original data
1	Corrected data
2	Reconstructed (such as by interpolation, estimation or disaggregation)
3	Calculated value

Table 3.2. Example of a stage of validation flag code

<i>Validation stage code</i>	<i>Meaning</i>
1	Missing data (data not received or observation not made)
2	Data eliminated once controls completed
3	Not controlled (newly inserted data or historical data not subject to any control)
4	Declared doubtful as identified as an outlier by preliminary checks, awaiting controls (data possibly wrong)
5	Declared doubtful after automatic controls or human supervision (data probably wrong)
6	Declared validated after automatic controls or human supervision (but further modification allowed, for example, if a subsequent study reveals that the data can still be improved)
7	Declared validated after automatic controls and human supervision and no further modification allowed

Table 3.3. Example of a data acquisition method flag code

<i>Acquisition method code</i>	<i>Meaning</i>
1	Global Telecommunication System
2	Key entry
3	Automated weather station telecommunication network
4	Automated weather station digital file
5	Manuscript record

These are examples of the many different ways to construct the quality information associated with data. An international data quality coding system is required to standardize quality information. WMO and its Services Commission is tackling this challenge.

2.4.6 Data errors

Data errors arise primarily as a result of instrumental, observer, data transmission, key entry and data validation process errors, as well as changing data formats and data summarization problems. When establishing a set of quality control procedures, all potential causes of error should be considered and efforts should be made to reduce them. It is recommended that, in developing automated and semi-automated error-flagging procedures, system designers work closely with operational quality control personnel.

Metadata errors often manifest themselves as data errors. For example, an incorrect station identifier may mean that data from one location apparently came from another; an incorrect date stamp may mean the data appear to have been observed at a different time. Data that are missing for the correct place and time should be detected by completeness tests; data that have been ascribed to an incorrect place or time should be detected by consistency and tolerance tests.

Format errors include repeated observations or impossible format codes such as alpha characters in a numeric field, embedded or blank fields within an observation, impossible identification codes, and impossible dates. The actual causes of a format error could include miskeying, the garbling of a message in transmission, or a mistake by an operator. Procedures should be introduced to eliminate, or at least reduce, format errors.

Completeness errors concern missing data. Incomplete data are much more critical for some elements than for others. For monthly extremes or event data such as the number of days with precipitation greater than a certain threshold, missing daily data may render the recorded value highly questionable. Total monthly rainfall amounts may also be strongly compromised by a few days of missing data, particularly when a rain event occurred during the missing period. On the other hand, monthly averaged temperature may be less susceptible to missing data than the two previous examples. For some applications data completeness is a necessity; since very few station datasets have 100% completeness over an extended period of time, estimating the missing data will normally require provisions for spatial or temporal interpolation.

Internal consistency relies on the physical relationships among climatological elements. All elements should be thoroughly verified against any associated elements within each observation. For example, psychrometric data should be checked to ensure that the reported dry bulb temperature equals or exceeds the reported wet bulb temperature. Similarly, the relationship between visibility and present weather should be checked for adherence to standard observation practices.

Data should be checked for consistency with definitions. For example, a maximum value must be equal to or higher than a minimum value. Physical bounds provide rules for further internal consistency checks. For example, sunshine duration is limited by the duration of the day, global radiation cannot be greater than the irradiance at the top of the atmosphere, wind direction must be between 0° and 360°, and precipitation cannot be negative.

Temporal consistency tests the variation of an element in time. Many climatological datasets show significant serial correlation. A check should be made by comparing the prior and subsequent observations with the one in question. Using experience or analytical or statistical methodologies, data reviewers can establish the amount of change that might be expected in a particular element in any time interval. For some elements, a lack of change over an extended period could indicate an error.

Spatial consistency compares each observation with observations taken at the same time at other stations in the area. Each observation can be compared to what would be expected at that site based on the observations from neighbouring stations. Those data for which there is a significant difference between the expected and actual observations should be flagged for review, correction or deletion as necessary. It is important to recognize that only like quantities should be directly compared, such as wind speeds measured at the same height; values measured at similar elevations, such as flat, open topography; or values measured within a climatologically similar area. Unlike quantities may sometimes be transformed or standardized to enable valid comparisons.

Summarization consistency errors can be detected by comparing different summaries of data. For example, the sums and means of daily values can be calculated for various periods such as weeks, months or years. Checking that the total of the twelve monthly reported sums equals the sum of the individual daily values for a year provides a quick and simple cross-check for an accumulation element such as rainfall. Systematic errors in upper-air station data can sometimes be revealed by comparing monthly averages with the averages derived for the same location and height from a numerical analysis system. Modelled data such as forecasts and reanalyses also can be used to detect consistency errors. The cause of any inconsistencies should be reviewed and corrected.

Tolerance errors are values that are impossible for an element, such as negative precipitation or, in other cases, where a value is unlikely. In the latter case, the limits are usually time- and location-dependent and should be established by recourse to the historical values or by spatial interpolation methods. It may be possible to perform some tolerance tests using completely different data streams, such as satellite or radar data. For example, a very simple test for the occurrence or non-occurrence of precipitation using satellite data would be to check for the presence of clouds in a satellite image.

2.4.7 Data rescue

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. It

includes organizing and imaging paper, microfilm and microfiche records; keying numerical and textual data and digitizing strip-chart data into a usable format; and archiving data, metadata and quality-control outcomes and procedures.

Guidelines on Best Practices for Climate Data Rescue ([WMO-No. 1182](#)) provides detailed guidance for all facets of climate data rescue. The International Data Rescue ([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 need 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. 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. It 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.

2.4.8 Exchange of climatic data

Exchange of data is essential for climatology. For WMO Members, the obligation to share data and metadata with other Members, and the conditions under which these may be passed to third parties, have been covered in the recent past under Resolution 60 of the Seventeenth World Meteorological Congress (with regard to the Global Framework for Climate Services), Resolution 25 of the Thirteenth World Meteorological Congress (for hydrological data), Resolution 40 of the Twelfth World Meteorological Congress (with regard to meteorological data), and Intergovernmental Oceanographic Commission Resolution XXII-6 (for oceanographic data). Resolution 1 of the 2021 Extraordinary World Meteorological Congress is one of three initiatives to meet the explosive growth in demand for weather and climate data products and services from all sectors of society. The demand is for weather, climate and water monitoring and prediction data to support essential services needed by all sectors of society.

The initiative under Resolution 1 is a [Unified Data Policy](#) that commits the WMO to broadening and enhancing the free and unrestricted international exchange of data from all parts of the world and of other data products among all WMO Members. A two-tiered approach to the international provision and exchange of Earth system data is maintained such that Members shall provide on a free and unrestricted basis the core data that are necessary for the provision of services in support of the protection of life and property and for the well-being of all nations, and that Members should also provide the recommended data that are required to support Earth system monitoring and prediction activities at the global, regional and national levels and to further assist other Members with the provision of weather, climate, water and related environmental services in their States and Territories. Conditions. Members should provide without charge access to all recommended data exchanged under the auspices of WMO to public research and education communities, for their non-commercial activities. The Resolution provides additional details describing data, elements and further recommendations.

The Unified Data Policy provides a comprehensive update of the policies guiding the international exchange of weather, climate and related Earth system data between the 193 Member states and territories of WMO. The new policy reaffirms the commitment to the free and unrestricted exchange of data, which has been the bedrock of WMO since it was established more than 70 years ago. Along with the other two initiatives, the [Global Basic Observing Network](#) and the [Systematic Observations Financing](#)

Facility, the WMO is dramatically strengthening the world's weather and climate services through a systematic increase in much-needed observational data and data products from across the globe.

The World Data System (WDS) is an Interdisciplinary Body of the [International Science Council](#). Its mission is to promote long-term stewardship of, and universal and equitable access to, quality-assured scientific data and data services, products, and information across all disciplines in the natural and social sciences, and the humanities. WDS aims to facilitate scientific research by coordinating and supporting the provision, use and preservation of relevant datasets, while strengthening their links with the research community. WDSs that collaborate with WMO are for meteorology and climatology, oceanography, paleoclimatology, solar radiation and radiation balance, glaciology, greenhouse gases, ozone and ultraviolet radiation, aerosols, and remote sensing (physics and chemistry). There is also collaboration in a publishing network for geoscientific and environmental data.

2.5 STEWARDSHIP MATURITY ASSESSMENT

[WMO-No. 1238](#) defines data stewardship as the formal accountability for ensuring effective controls around the management and use of the climate record. Stewardship assigns rights and responsibilities for acquiring and managing climate data and information. The rights include the determination of how the information will be managed and any access constraints, with accompanying responsibilities towards maintenance, quality, security and enabling appropriate access to that information. A unified framework for measuring the level of stewardship practices applied to data is a stewardship maturity matrix which generally defines measurable, level-progressive practices of key components of stewardship such as preservation, accessibility, and transparency/traceability, rating each component on a level scale from not managed to optimally managed.

A stewardship maturity assessment in the form of a matrix measures how well a dataset has been created and curated to ensure the accessibility, usability and integrity of the data, and sufficient documentation for data users. It will necessarily be limited to those facets which can be (independently) assessed. The assessment provides information on the extent to which a dataset has clear documentation, support channels, is constructed with clear coding practices, applies quality control and assurance procedures, provides uncertainties and adheres to data format and archiving standards. However, it does not explicitly assess the scientific rigour involved in creating the dataset, that is, how reliable the underpinning observations are, details of processing, homogenisation, scientifically-based adjustments, and so forth.

Once the stewardship maturity of a dataset has been assessed and scores are available, there are a number of different ways both data managers and users of the datasets can use this information. For data managers, having an independent set of assessments across a number of aspects could be useful in identifying where to focus limited resources in improving stewardship quality. There may be some aspects that require little effort to obtain higher ratings. Furthermore, by contrasting ratings against other similarly well-managed products, the scores may even help prioritize cost planning, resource allocation and funding for future data management with the aim of improving stewardship maturity for those datasets. Dataset creators can use the scores similarly when outlining major updates or ensuring stewardship maturity of new datasets.

There are a number of ways data users can or should use the scores from this matrix. At a simple and high level for users with minimal requirements the scores can be used to choose the dataset with the

highest level of maturity for their specific application. Mature datasets and systems make it easy for users to assess which dataset they need. However, it is highly encouraged that users take a more in-depth approach, thinking about their application as well as the scores for each category and aspect. Datasets which have different aims and processing levels will have different maturity scores, but the appropriate dataset for a particular user's application may be one with a lower overall score. For example, when studying sea surface temperatures, a user could access a highly-processed gridded dataset which has not been assessed. Alternatively, one could use a dataset of raw ship track and buoy information which has been assessed. However, just because this dataset has been assessed does not necessarily make it the right choice for a particular application. Furthermore, a higher assessed dataset score does not automatically make it the right choice.

More detailed information about the rationale, development and uses of stewardship maturity matrices and assessments is given in [Dunn et al. \(2021\)](#).

2.6 REFERENCES

2.6.1 Cited references

Dunn, R., C. Lief, G. Peng, W. Wright, O. Baddour, M. Donat, B. Dubuisson, J.-F. Legeais, P. Siegmund, R. Silveira, X.L. Wang and M. Ziese, 2021: [Stewardship Maturity Assessment Tools for Modernization of Climate Data Management](#). *Data Science Journal*, 20(1), p.7. <http://doi.org/10.5334/dsj-2021-007>.

World Meteorological Organization, 1983, 1996, 2008, 2014: *Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8)*. Fifth, sixth, seventh and eighth editions. Geneva.

———, 1986: *Guidelines on the Quality Control of Surface Climatological Data (WMO/TD-No. 111)*. Geneva.

———, 1997: *Report of the GCOS/GOOS/GTOS Joint Data and Information Management Panel (WMO/TD-No. 847)*

———, 2003: *Guidelines on Climate Observation, Network and Systems (WMO/TD-No. 1185, WCDMP-No. 52)*. Geneva.

———, 2003: *Guidelines on Climate Metadata and Homogenization (WMO/TD-No. 1186, WCDMP-No. 53)*. Geneva.

———, 2006: *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites (WMO/TD-No. 1250)*. Geneva.

———, 2008: *Technical Regulations (WMO-No. 49)*, Volume III: Hydrology. Geneva.

———, 2009: *Guide to Hydrological Practices (WMO-No. 168)*, Volume II: Management of Water Resources and Application of Hydrological Practices. Geneva.

———, 2009: *Plant Phenological Guidelines (WMO/TD-No. 1484, WCDMP-No. 70)*. Geneva.

———, 2012: *Guide to Agricultural Meteorological Practices. (WMO-No. 134)*. Geneva.

———, 2013: *The GCOS Reference Upper-Air Network (GRUAN): Guide (GCOS-No. 171)*. Geneva.

- , 2013: *Fourth session of the Expert Team on Marine Climatology (ETMC): Final Report* ([JCOMM Meeting Report No. 94](#)). Geneva.
- , 2014: *Climate Data Management System Specifications* ([WMO-No. 1131](#)). Geneva.
- , 2016: *The Global Observing System for Climate: Implementation Needs* ([GCOS–No. 200](#)). Geneva.
- , 2016: *WMO/GAW Aerosol Measurement Procedures, Guidelines and Recommendations* ([WMO-No. 1177](#), GAW-No. 227). Geneva.
- , 2016: *Guidelines for Best Practices for Climate Data Rescue* ([WMO-No. 1182](#)). Geneva.
- , 2017: *Guide to the Global Observing System* ([WMO-No. 488](#)). Geneva.
- , 2017: *Guidelines on Best Practices for Achieving User Readiness for New Meteorological Satellites* ([WMO-No. 1187](#)). Geneva.
- , 2018: *Guide to Meteorological Instruments and Methods of Observation* ([WMO-No. 8](#)). Geneva.
- , 2018: *Manual on Marine Meteorological Services* ([WMO-No. 558](#)). Volume I: Global Aspects. Geneva.
- , 2018: *Guide to Competency* ([WMO No. 1205](#)). Geneva.
- , 2019: *Manual on the Global Data-processing and Forecasting System* ([WMO-No. 485](#)). Geneva.
- , 2019: *Manual on the WMO Information System* ([WMO-No. 1060](#)). Geneva.
- , 2019: *Manual on the WMO Integrated Global Observing System* ([WMO-No. 1160](#)). Geneva.
- , 2019: *Guide to the WMO Integrated Global Observing System* ([WMO-No. 1165](#)). Geneva.
- , 2019: *WIGOS Metadata Standard* ([WMO-No. 1192](#)). Geneva.
- , 2019: *Compendium of WMO Competency Frameworks* ([WMO- No. 1209](#)). Geneva.
- , 2019: *The High-quality Global Data Management Framework for Climate* ([WMO-No. 1238](#)). Geneva.
- , 2020: *Guide to Hydrological Practices* ([WMO-No. 168](#)), Volume I: Hydrology – From Measurement to Hydrological Information. Geneva.
- , 2021: *Technical Regulations* ([WMO-No.49](#)), Volume I: General Meteorological Standards and Recommended Practices. Geneva.
- , 2021: *Technical Regulations* ([WMO-No. 49](#)), Volume II: Meteorological Service for International Air Navigation. Geneva.
- , 2021: *Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications* ([WMO-No 1269](#)).

2.6.2 Additional reading

Boubakeur, G., 2021: [Climate as the major factor controlling phenology](#). In: *Agrometeorology* (R.S. Meena, ed.), London, IntechOpen, <https://doi.org/10.5772/intechopen.95893>.

Cornford, D., 1998: An overview of interpolation. In: *Seminar on Data Spatial Distribution in Meteorology and Climatology* (M. Bindi and B. Gozzini, eds.). Volterra, European Union COST Action 79.

Cronin, T.M., 1999: *Principles of Paleoclimatology*. New York, Columbia University Press.

De Gaetano, A.T., 1997: [A quality control procedure for hourly wind data](#). *Journal of Atmospheric and Oceanic Technology*, 14:137–151, [https://doi.org/10.1175/1520-0426\(1997\)014%3C0308:AQCRFH%3E2.0.CO;2](https://doi.org/10.1175/1520-0426(1997)014%3C0308:AQCRFH%3E2.0.CO;2).

Dunn, R., c. Lief, G. Peng, W. Wright, O. Baddour, M. Donat, B. Dubuisson, J.-F. Legeais, P. Siegmund, R. Silveira, X.L. Wang and M. Ziese, 2021: [Stewardship maturity assessment tools for modernization of climate data management](#). *Data Science Journal*, 20(1), p.7, <http://doi.org/10.5334/dsj-2021-007>.

Free, M., I. Durre, E. Aguilar, D. Seidel, T.C. Peterson, R.E. Eskridge, J.K. Luers, D. Parker, M. Gordon, J. Lanzante, S. Klein, J. Christy, S. Schroeder, B. Soden and L.M. McMillin, 2002: [Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring](#). *Bulletin of the American Meteorological Society*, 83:891–899, [https://doi.org/10.1175/1520-0477\(2002\)083%3C0891:CCRDCW%3E2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083%3C0891:CCRDCW%3E2.3.CO;2).

Free, M. and D.J. Seidel, 2005: [Causes of differing temperature trends in radiosonde upper air datasets](#). *Journal of Geophysical Research*, <https://doi.org/10.1029/2004JD005481>.

Graybeal, D.Y., A.T. De Gaetano and K.L. Eggleston, 2004: [Complex quality assurance of historical hourly surface airways meteorological data](#). *Journal of Atmospheric and Oceanic Technology*, 21:1156–1169, [https://doi.org/10.1175/1520-0426\(2004\)021%3C1156:CQAOHH%3E2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021%3C1156:CQAOHH%3E2.0.CO;2).

Hsu, A., H., W. Khoo, N. Goyal and M. Wainstein, 2020: [Next-generation digital ecosystem for climate data mining and knowledge discovery: A review of digital data collection technologies](#). *Frontiers in Big Data*. 3:29, <https://www.frontiersin.org/articles/10.3389/fdata.2020.00029/full>.

International Organization for Standardization, 2015: *Quality management systems – Fundamentals and vocabulary*, ISO 9000:2015. Geneva.

———, 2015: *Conformity Assessment – Requirements for Bodies Providing Audit and Certification of Management Systems*, ISO/IEC 17021-1:2015. Geneva.

———, 2018: *Quality Management – Quality of an Organization – Guidance to Achieve Sustained Success*, ISO 9004:2018. Geneva.

Karl, T.R., C.N. Williams Jr, P.J. Young and W.M. Wendland, 1986: [A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States](#). *Journal of Climate and Applied Meteorology*, 25:145–160, [https://doi.org/10.1175/1520-0450\(1986\)025%3C0145:AMTETT%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025%3C0145:AMTETT%3E2.0.CO;2).

Kim, H. J., S.M. Park, B.J. Choi, S.H. Moon and Y.H. Kim, 2020: [Spatiotemporal approaches for quality control and error correction of atmospheric data through machine learning](#). *Computational Intelligence and Neuroscience*. 7980434, <https://doi.org/10.1155/2020/7980434>.

Kresse, W. and K. Fadaie, 2004: *ISO Standards for Geographic Information*. Berlin, Springer.

Leroy, M., 1998: Meteorological measurements representativity: nearby obstacles influence. In: *Tenth Symposium on Meteorological Observations and Instrumentation*, Phoenix, Arizona, 11– 16 January 1998. Boston, American Meteorological Society.

— — —, 1998: Climatological site classification. In: *Tenth Symposium on Meteorological Observations and Instrumentation*, Phoenix, Arizona, 11–16 January 1998. Boston, American Meteorological Society.

Lieth, H. (ed.), 1974: *Phenology and Seasonality Modeling*. New York, Springer.

Martin, D.J., A. Howard, R. Hutchinson, S. McGree and D.A. Jones, 2015: [Development and implementation of a climate data management system for western Pacific small island developing states](#). *Meteorological Applications*, 22(2): 273-287, <https://doi.org/10.1002/met.1461>.

Nakagawa, Y., Y. Onoue, S. Kawahara et al., 2020: [Development of a system for efficient content-based retrieval to analyze large volumes of climate data](#). *Progress in Earth and Planetary Science*, 7(9), <https://doi.org/10.1186/s40645-019-0315-9>.

Napoly, A., T. Grassmann, F. Meier and D. Fenner, 2018: [Development and application of a statistically-based quality control for crowdsourced air temperature data](#). *Frontiers in Earth Science*, <https://doi.org/10.3389/feart.2018.00118>.

Nightingale J, K.F. Boersma, J.-P. Muller, S. Compennolle, J.-C. Lambert, S. Blessing, R. Giering, N. Gobron, I. De Smedt, P. Coheur, M. George, J. Schulz and A. Wood, 2018: [Quality assurance framework development based on six new ECV data products to enhance user confidence for climate applications](#). *Remote Sensing*, 10(8):1254, <https://www.mdpi.com/2072-4292/10/8/1254#>.

Oke, T.R., 1987: *Boundary Layer Climates*. Second edition. Cambridge, Cambridge University Press.

Schwartz, M.D. (ed.), 2003: *Phenology: An Integrative Environmental Science. Tasks for Vegetation Science*, Volume 39. Dordrecht, Kluwer Academic Publishers.

Seidel, D.J. and M. Free, 2006: [Measurement requirements for climate monitoring of upper air temperature derived from reanalysis data](#). *Journal of Climate*, 19:854–871, <https://doi.org/10.1175/JCLI3666.1>.

Singleton, F. and E.A. Spackman, 1984: Climatological network design. *The Meteorological Magazine*, 113:77–89.

Sneyers, R., 1973: Sur la densité optimale des réseaux météorologiques. *Archiv für Meteorologie, Geophysik und Bioklimatologie*, Serie B, 21:17–24.

Thorne, P.W., D.E. Parker, J.R. Christy and C.A. Mears, 2005: [Uncertainties in climate trends: Lessons from upper air temperature records](#). *Bulletin of the American Meteorological Society*, 86:1437–1444, <https://doi.org/10.1175/BAMS-86-10-1437>.

Wang, J., D.J. Carlson, D.B. Parsons, T.F.Hock, D. Lauritsen, H.L. Cole, K. Beierle and E. Chamberlain, 2003: [Performance of operational radiosonde humidity sensors in direct comparison with a chilled mirror dew-point hygrometer and its climate implication](#). *Geophysical Research Letters*, 30:10.1029/2003GL016985, <https://doi.org/10.1029/2003GL016985>.