



WMO Commissions. Expert Teams have been established by CCI for metadata and for data rescue. Implementation Coordination Teams within CCI have been tasked with ensuring that the work of the Expert Teams is communicated to, and implemented within, all the WMO Regions. Activities such as the Climate Computing project (CLICOM) and the Climate Database Management System (CDMS) project help provide the sound database management infrastructure required by WMO Members.

This article aims to highlight the important issues concerned with climate data and data management, what changes are occurring and how WMO is supporting its Members in the provision of climate services that will help them to exploit their natural resource base at sustainable levels and prepare for the climate of the future.

### Climate observation networks and systems

Observations of the atmosphere, ocean and land, such as those taken for routine weather forecasting, are important because they help satisfy essential social, economic and environmental needs and, furthermore, they are an integral part of national infrastructure to reduce the risk of loss of life and damage to property. Senior managers in NMHSs are encouraged to remind their governments regularly of these reasons for recording, collecting and managing such observations.

Climate observations have traditionally encompassed the many meteorological and related observations used in applications such as weather forecasting, air pollution modelling and environmental impact assessments. However, it is now recognized that an understanding of climate variability and change requires information about all three domains of the Earth system, i.e. atmosphere, ocean and land, including the Earth's radiation balance. Commensurate with this recognition is the need for a comprehensive set of observing systems. The observations collected thus far remain invaluable, but new techniques and methods are required to ensure that operational climate services and researchers can provide answers to the problems that beset us now and that will do so in the future.

As a consequence of climate being more than simply the "average" of weather observed over time, an adequate climate database will be more complex than the sum of all "weather" observations. Firstly, as noted, climate observations need to account for the full range of elements that describe the state and variability of the climate system—not just those that describe the atmosphere. Extensive observations of the ocean and

terrestrial-based systems are also required. Secondly, an observation at any point in time needs a reference climate against which it can be judged, i.e. there must be some way of deciding upon a reference climatological period. In this regard, observations from a station that exists for only a short period (i.e. from days to a few years) or which relocates very frequently will generally be of less value than those observations from a station whose records have been maintained to set standards over many years. Thus, in order to derive a satisfactory climatological average (or normal) for a particular climate element, a long-period record of homogeneous, continuous and good-quality observations for that element is required. This has obvious implications for the design and operation of observational networks and systems. Thirdly, a climate observation should be associated—either directly or indirectly—with a set of metadata that will provide users with information, often implicitly, on how the observation should be interpreted and used.

While climate observations will serve multiple purposes beyond the need to describe climate itself, it is essential that they have the particular characteristics that serve this basic requirement. Increasingly, climate observations are being derived from satellite-borne instruments, which makes the maintenance of robust surface-based systems against which the new data can be referenced and compared even more critical. Further, we are beginning to see new systems being deployed in the world's oceans that are adding to the challenge of maintaining a comprehensive record of the Earth's climate. A strong case for improved climate observations has been made by Karl *et al.* (1995), and Trenberth *et al.* (2002).

Climate observations will attain their true value only if the data, and the information that derive from them, are used. This means that the data must be readily accessible and transformable into products and services and used in applications. The basic monthly, seasonal and annual summaries of temperature, rainfall and other climate elements provide an essential resource for planning endeavours in areas such as agriculture, water resources, emergency management, urban design, insurance, energy usage and construction. Climate data are also unlocking important relationships between climate and health, including the effects of extreme heat and cold on mortality. Millions of people each year use climatological information in planning their annual vacations. In a relatively new area of applications, high-quality climate observations are being used by the weather derivatives industry, which has already traded billions of dollars based to a large degree on climate information.

The need for more accurate detection and analysis of climate change and the promise of further improvements in seasonal-to-interannual (SI) prediction have further increased the value of climate data in recent decades. Climate data are fundamental to the operation and validation of climate models, which are widely used for SI prediction and for generating projections of future climate. Maximizing the availability of computerized historical data, including metadata, is paramount for long-term climate monitoring—particularly for analysing trends in the occurrence of extreme events where data quality and longevity of record become even more significant considerations (Karl and Easterling 1999; Nicholls 1995).

The complexity of demands on the systems for recording and monitoring climate, including the detection of climate change, has led to the development of special networks at national (e.g. Reference Climate Stations), regional (e.g. Regional Basic Climatological Network) and global (e.g. Global Climate Observing System (GCOS) Surface Network) scales. Hence, when designing and managing observational networks, NMHSs must consider the purposes and requirements of the observations at different scales.

As previously discussed, there are some important characteristics that meteorological and related observations require for them to be classed as fully acceptable for climate purposes. Karl *et al.* (1995) and NRC (1999) have laid down and described some guiding principles for long-term sustainable climate monitoring. While these principles were primarily identified for the purposes of improving our ability to detect climate change, they are widely applicable to all facets of climate observations. The 10 principles as adopted by the United Nations Framework Convention on Climate Change (UNFCCC) and endorsed by CCL XIII are the following:

- The impact of new systems or changes to existing systems should be assessed prior to implementation;
- A suitable period of overlap for new and old observing systems is required (Figure 2);
- The details and history of local conditions, instruments, operating procedures, data processing algorithms and other factors pertinent to interpreting data (i.e. metadata) should be

documented and treated with the same care as the data themselves;

- The quality and homogeneity of data should be regularly assessed as a part of routine operations;
- Consideration of the needs for environmental and climate-monitoring products and assessments, such as Intergovernmental Panel on Climate Change (IPCC) assessments, should be integrated into national, regional and global observing priorities;
- Operation of historically uninterrupted stations and observing systems should be maintained;
- High priority for additional observations should be focused on data-poor regions, poorly-observed parameters, regions sensitive to change, and key measurements with inadequate temporal resolution;
- Long-term requirements, including appropriate sampling frequencies, should be specified to network designers, operators and instrument engineers at the outset of system design and implementation;
- The conversion of research observing systems to long-term operations in a carefully-planned manner should be promoted; and
- Data-management systems that facilitate access, use and interpretation of data and products should be included as essential elements of climate monitoring systems.

Much can, and has, been said about these principles but there are perhaps two basic messages for climatologists. Firstly, climatologists cannot afford to be passive when it comes to accepting proposed observation network and system changes. They must open



Figure 2 — An example of a parallel observations programme in Helsinki, Finland. Such programmes are required when there are changes in observation systems or site location and comparison data are required to enable homogeneity to be preserved. (Source: Finnish Meteorological Institute, Finland)

dialogue and build relationships with managers of those networks and systems, and also with data-communication managers and database managers so that specific climate needs are understood and appreciated. Secondly, climatologists must accept that there will be changes in observation networks and systems, regardless of how much they would like them to remain the same. The key issue is to ensure that such changes are managed. This will require, amongst other things, the establishment of robust, well-promulgated change-management practices within the responsible organisation (T. Allsopp, personal communication, 2002).

As a result of a recent meeting in Malaga, Spain (24-26 February 2003) (see page 305), the WCDMP—through the CCI—began to develop guidance material on observation networks and systems (as well as for data rescue, and for metadata homogenization). It is expected that this initiative will trigger the development of further guidance material on climate data management, including quality control. In addition, CCI experts are involved in framing these requirements within the context of the Global Climate Observing System (GCOS) and also with regard to the statements of guidance on observation requirements for all major users of atmospheric, oceanographic and terrestrial observations. WMO is actively engaging in and supporting relevant conferences and workshops on observation networks and systems and, in particular, is promoting the implementation of the principles for long-term climate monitoring.

There are significant challenges ahead for WMO and all its Members in improving climate observation networks and systems. The second adequacy report on The Global Observing Systems for Climate, coordinated by the GCOS Secretariat, concludes that existing atmospheric, oceanographic and terrestrial observation systems have a long way to go before they perform adequately to the standards required of climate.

## Metadata

The word metadata is a combination of the Greek *meta* (beyond) and the plural of the Latin *datum* (a given fact). Metadata—i.e. data about data—must reflect how, where, when and by whom the observational data were collected and what happened between the moment they were observed and the moment they were archived.

Metadata are sometimes termed as “directory level” when they refer to information about the contents of datasets, and “archive level” when reference is made to information about the recording and collection process, i.e. what climatologists sometimes term “station history” information. Here, we focus primarily on the latter. Ideally, a complete metadata record should keep track of any changes associated with a station’s operating environment that could affect its output data (Figure 3).

Good-quality metadata will provide information about the conditions under which observational data have been recorded, gathered and transmitted and thus enable a user of the data to draw appropriate and valid conclusions. Conversely, the lack of metadata may lead to misinterpretations. Climatologists must inform and educate those involved in data-collection and management processes about metadata requirements and why they are important. Only then can we be confident of the completeness, accuracy and relevance of the records themselves to the task at hand. What might be an obvious and common practice at some point in time, may lead to a future source of uncertainty or error if it is not well documented. As an example, it is common in many countries to encode a trace amount of precipitation with “-3” or other negative value. This is a very convenient “flagging” strategy, because a precipitation total can never be lower than zero. However, in the absence of metadata, some users may be inclined to ignore a value below zero for this element, which would then lead to an underestimation of the number of rainy days. If observers and/or data managers have been careful enough to document this practice, and the user takes time to understand the information before doing any analysis, such mistakes should not happen. As a further example, the documenting of the exact date and

### What are metadata and why do we need them?

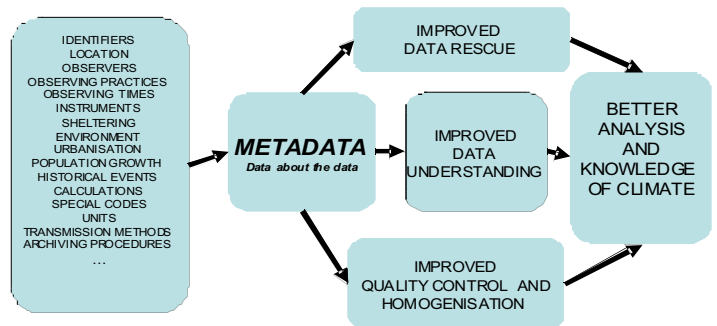


Figure 3 — The data sources for metadata are many and varied but with careful recording, rescue efforts and sound management, there are substantial benefits for climate analysis and, ultimately, for improving climate knowledge.

time when an instrument is replaced and of the technical characteristics of both the new and the old instruments will help to remove non-climatic “fingerprints” associated with such changes on climate records. Similarly, the details of station moves and changes in the characteristics of an ongoing site, e.g. the erection of buildings and vegetation growth, are essential for data users, particularly in assessing long-term climate variations.

In a little more than a decade, the IPCC has raised the awareness of global warming to the point that it is now one of the most significant issues for humanity. Metadata play a key role in the creation of the high quality and homogeneous long-term datasets required to meet the IPCC’s exacting standards as well as those necessary for operational climate services..

Meteorological data users working in fields such as agricultural meteorology, engineering or aeronautics, also benefit from good metadata. Since these users often compare data derived from many locations and sensors at different times, they too need to ensure that the data are reliable and can be compared. Information about different observing conditions and practices at different sites will help them to achieve this goal.

As will be discussed below, WMO is putting great efforts into data rescue projects, since it is clear that many valuable meteorological data remain in non-digital form and are at risk of being lost. Such data rescue campaigns would also save the associated metadata, together with the observational data. In this regard, independent efforts may be required for the metadata, which are often not collocated with the recorded observational data. When gathering metadata on past observations, it is advisable to add information about historical events that may affect the quality, homogeneity and integrity of the record. Events such as the creation or disappearance of national boundaries, wars, population growth, creation of institutions, establishment of new policies, etc., may all have an impact on an individual station or indeed on a whole network and will therefore be relevant. data rescue projects also benefit from the availability of metadata. For example, comparing the available digitized records from a set of stations with metadata that details their operating periods gives an indication of the potential benefits to be gained from a data rescue effort.

The previously mentioned CCI meeting in Malaga on producing guidance material also drew together experts to draft WCDMP guidance on metadata practices. The purpose of this guide is to provide

NMHSs with information on what metadata are important and how they should be managed. The guide will include a comprehensive list of the essential items to be recorded. For example, a good metadata record for a meteorological station should include all the information needed for identification and location; national and international codes; name and aliases; geographical information (latitude, longitude and elevation); contact details (responsible institution, observer, etc.), and period of operation (starting date, interruptions and closing date). It is also important to record accurate information about the instruments in use, sheltering, maintenance and possible obstacles, as well as the characteristics of the local environment of the station and land use. The knowledge of observation codes, units, observing times, measured elements, corrections applied to data and the method used to calculate derived values (e.g. daily averages) are also extremely useful in avoiding misinterpretations, and should be included in a complete metadata record. It is also advisable to indicate any quality control or homogenization adjustments that have been made to data prior to transmission. Other sources of valuable metadata are histories of the observation of individual parameters, including observational practices for observed elements (e.g. temperature, rainfall). This information will generally be represented in the form of historical summaries extracted from individual station histories. Quality control should be applied to the metadata as well—and preferably soon after the information has been recorded.

WMO has a strong interest in encouraging and supporting the recording and management of metadata and metadata recovery efforts. In terms of their management, the same standards and practices that apply to observational data generally apply to metadata, i.e. data should be secure, accessible and subject to quality control, etc. While many metadata should be stored ideally within a modern climate database system (e.g. a relational database), it is recognized that this may not yet be feasible for many countries and, indeed, not practical for all sources of metadata. The CCI Expert Team on Metadata for Climate Applications will be working towards developing solutions that are relevant for all NMHSs.

### **Data rescue, preservation and digitization of climate data**

An NMHS that routinely produces high-quality observations, records and manages its metadata and maintains a modern, robust and secure database, may still be “holding back supply” if it has a stockpile of paper

records awaiting digitization (Figure 4). Unfortunately this situation is commonplace and many countries around the world are unable to take advantage of the full potential of their climate records.

The practice of conscientiously recording weather observations and transcribing them onto paper forms has been at the heart of climatology since its inception. The entering of data into a computer medium has become a more common practice in most countries in more recent decades. However, due to a lack of funding, lack of expertise, or the demands of keeping up with new data flows, many invaluable past data have never made the leap from paper to computer. An unfortunate consequence is that data stored on millions of paper forms are now at risk of being lost due to the physical deterioration of the paper. To some extent, similar comments apply to data stored on other media such as microfilm, microfiche and older types of computer media, e.g. magnetic tapes. For some countries, their records are spread among several institutions (and sometimes even located in other countries!).

6

Urgent action is required to ensure that the vast amount of climate data collected are properly preserved in an easily accessible and useful form. At a recent International Data Rescue Meeting (WMO 2002), data rescue was defined as being “an ongoing process of preserving all data at risk of being lost due to deterioration of the medium, and the digitizing of current and past data into computer compatible form for easy access”. It was added that:

- Data should be stored as image files onto media that can be regularly renewed to prevent the

deterioration of the medium (cartridges, CDs, DVDs, etc.);

- Data already in computer-compatible media should be constantly migrated to storage facilities that conform to changing technologies; and
- Data should be key-entered in a form that can be used for analyses.

The essential steps in a data rescue effort are shown in Figure 5.

The establishment of the Data Rescue project (DARE) as a high priority within WMO is critically important if we are to ensure that future generations of scientists and other data users have access to all information necessary for their research, services and applications. WMO has had several successes with DARE projects undertaken in several parts of the world. For example, a project funded primarily by Belgium resulted in millions of documents from about 50 African countries being saved on microfiche and microfilm. The joint CCI/CLIVAR Working Group on Climate Change Detection conducted a workshop on climate extremes in the Caribbean (Peterson *et al.* 2002) that demonstrated what can be achieved through regional collaboration and with relatively limited resources. Following this workshop, a number of countries increased their efforts to digitize records.

The case for long-period historical climate records and DARE activities was discussed earlier. Recent IPCC assessments on climate change indicate that long-period records of quality climate data are critical for improving the understanding of climate variations and climate change. The need for expanding the range and availability of digitized

data extends, however, beyond this basic need to describe climates of the past. For example, a key impetus for data rescue efforts in Australia has come from the agricultural community, which has recognized the importance of climate data for agricultural simulation models, for improved seasonal to interannual prediction and for better understanding of the effects of climate extremes on productivity (Clarkson *et al.*, 2001).

At the recent Malaga meeting, WMO CCI experts began drafting guidance material on data rescue. When completed, this guidance will provide NMHSs with essential information required to



Figure 4 — Historical climate records awaiting digitizing. It is important that—unlike this example—records are maintained in a suitably air conditioned room.

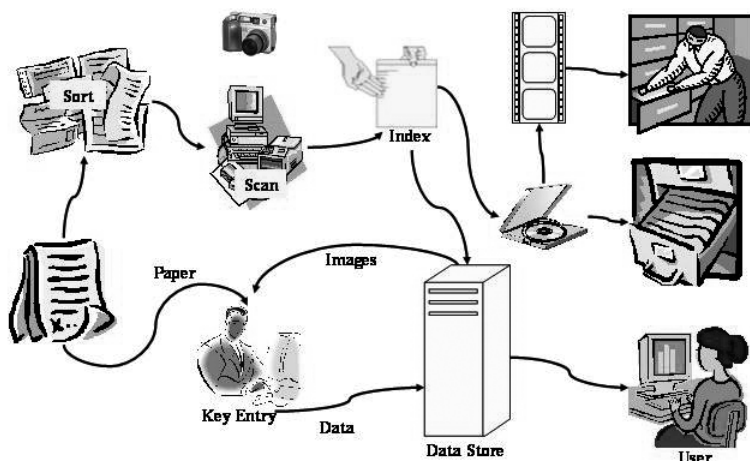


Figure 5 — Steps required in data rescue (Source: National Climate Centre, Bureau of Meteorology, Australia)

of the WMO CCI Task Group on future WMO CDMS (WMO, 2000). Some interesting results were obtained after the data were grouped into two broad categories of countries classified as “least developed” and “others”. There were some differences in functional requirements and priorities presented by these two groups of countries. Items ranked most important by the “least developed” countries were quality-control features, provision of user manuals and system documentation,

identify, preserve and digitize their historical climate records.

### Climate-database management systems

The CLICOM project started in the mid-1980s in order to meet the climate-data management needs of a considerable number of WMO Members which did not have the capability to use computers to manage their climate data. While several countries continue to provide substantial support to others in the implementation and maintenance of CLICOM systems, significant software components are no longer available and the evolution of hardware and operating systems has made it difficult to guarantee ongoing support.

WMO is in the process of coordinating the replacement of the current CLICOM system with Climate Database Management Systems (CDMSs). The WCDMP Division of WMO conducted a survey in November 1999 among Members, asking if their NMHSs had developed (or were developing) a CDMS that they would be prepared to make available for possible future use by other Members. All the specifications that might apply to a CDMS were set down in tables of functional requirements that had been prepared at a CDMS workshop (Toulouse, France, 30 March-1 April 1999). NMHSs were asked to use the tables to rank the different functional requirements and support services of future CDMSs. Those NMHSs that had candidate CDMSs were also asked to use the table to describe their CDMSs fully. Replies from 96 Members were received and 84 per cent expressed interest in using, testing or helping to enhance future CDMSs.

A detailed analysis of the 80 items in the tables of functional requirements was presented at the meeting

descriptions of database entity-relationship(E-R) model, security management, methods used to generate products in agricultural climatology, methods and facilities to homogenize data, and finally, production of CLIMAT messages. Items given the lowest priority were acquisition of AWS and SHIP data. Gridded data were given consistently low priority in all Regions except South America, where, presumably, large data gaps make interpolation more important. The meeting concluded that the “least developed” countries, as one of the main target groups for the new WMO CDMS project, should have their particular requirements strongly reflected in the CDMSs that are proposed for them.

The CDMS initiative is a significant step toward an integrated approach to data archiving and management. The special needs of Members and the functional requirements of the CDMS framed the evaluation criteria proposed for the testing and self evaluation of the CDMS (WMO 2000). The following were also significant in the project development:

- Practical experience in past CLICOM implementation and development;
- Regional CLICOM training seminars and workshops;
- A Meeting of the CCI Task Group in Ostrava, Czech Republic, in November 1998 (WMO, 1999) and its follow-up Workshop in Toulouse in April 1999; and
- An Expert Meeting in Toulouse in May 1997 (WMO, 1998).

Essential requirements expected from CDMSs included the following:

- Enhanced ability of Members to archive data in real-time or in delayed operation and to produce climatological summaries efficiently;

- Improved management of station metadata, including historical information of the stations, measuring equipment, inspection reports and site comments;
- Ensure that systems have the capacity to interface with data-entry systems to allow data to be digitized;
- Provision of a comprehensive quality-control and validation system; and
- Ability to handle data created by the CLICOM system.

Several countries have been actively involved in the testing and evaluation of their CDMSs since the CCI Task Group Meeting in May 2000. Two external experts were invited to evaluate and document the systems offered by WMO Members, in an evaluation workshop from 27 May to 1 June 2002 in Geneva. Considerable progress has been made in this exercise and, so far, six systems have been fully evaluated.

In the implementation of a new CDMS, it is important to understand the immense responsibilities involved in installation, customization, documentation, training and software maintenance, in addition to the costs incurred in purchasing systems. A successful implementation needs considerable commitment from NMHSs and an effective framework that facilitates the delivery of these requirements. The framework is expected to provide a direct and rapid assessment of the impacts of new requirements and the redirection of resources accordingly.

## Conclusion

Future climate services within NMHSs will need to be underpinned by a number of key capabilities. Climate observations will need to be taken from observation networks and systems that provide the high standards necessary for a range of climate needs, including climate change detection. The UNFCCC-agreed Climate Monitoring Principles will need to be adhered to for all observing networks designed for monitoring long-term trends and for other networks wherever possible—the benefits clearly extend beyond an understanding of climate itself. NMHSs will need to identify, recover and digitize their historical paper records, including metadata, which should be managed with the same care as the observational data. Robust and secure climate-database management systems will be necessary to serve the increasing number of climate products and services expected from NMHSs. NMHSs will also need to migrate their climate archives to newer media and database-management systems as

required, to ensure that they are not lost through obsolescence of hardware and operating systems. The World Climate Programme and the CCI will continue to serve NMHSs in helping them meet these needs.

## Acknowledgements

The authors are grateful for useful comments from several colleagues, particularly Dr Paul Llanso (Chief, WMO World Climate Data and Monitoring Programme), Dr Michael Coughlan, Ian Muirhead and John Shortridge (National Climate Centre, Bureau of Meteorology, Australia).

## References

- CLARKSON, N.M., B. TREWIN, D. JONES, N. PLUMMER, R. HUTCHINSON, and K. WONG, 2001: Extending the Computerized Australian Climate Archives to Unlock our Climate History—the CLIMARC Project. 14th Australia-New Zealand Climate Forum, Darwin, 18-21 September 2001, 49 pp.
- GCOS, 2003: Second Report on the Adequacy of the Global Observing System for Climate. Secretariat of the Global Climate Observing System (GCOS), Geneva. (In preparation).
- KARL, T.R., V.E. DERR, D.R. EASTERLING, C.K. FOLLAND, D.J. HOFFMAN, S. LEVITUS, N. NICHOLLS, D.E. PARKER and G.W. WITHEE, 1995: Critical issues for long-term climate monitoring. *Climatic Change*, 31, 185-221.
- KARL, T.R. and D.R. EASTERLING, 1999: Climate extremes: selected review and future research directions. *Climatic Change*, 42, 309-325.
- NICHOLLS, N. 1995: Long-term climate monitoring and extreme events. *Climatic Change*, 31, 231-245.
- National Research Council (USA), 1999: *Adequacy of Climate Observing Systems*. National Academy Press, USA, 51 pp.
- PETERSON, T.C., M.A. TAYLOR, R. DEMERITTE, D.L. DUNCOMBE, S. BURTON, F. THOMPSON, A. PORTER, M. MERCEDES, E. VILLEGAS, R. SEMEXNT FILS, A.K. TANK, A. MARTIS, R. WARNER, A. JOYETTE, W. MILLS, L. ALEXANDER and B. GLEASON, 2002: Recent changes in climate extremes in the Caribbean region. *Journal of Geophysical Research*, 107, D21, 4601.
- TRENBERTH, K.E., T.R. KARL, and T.W. SPENCE, 2002. The need for a systems approach to climate observations. *Bulletin of the American Meteorological Society*, 83, 1593-1602.
- WMO, 1998: Expert Meeting to review and assess the Oracle-based Prototype for Future Climate Database Management Systems (CDMS), Toulouse, 12-16 May 1997. WCDMP No. 34.
- WMO, 1999: Report of the Meeting of the WMO Commission for Climatology Task Group on Future WMO Climate Database Management Systems, Ostrava, Czech Republic, 10-13 November 1998. WCDMP No. 38.
- WMO, 2000: Meeting of the WMO Commission for Climatology Task Group on Future WMO Climate Database Management Systems, Geneva 3-5 May 2000. WCDMP No. 46.
- WMO, 2002: Reports of the CLICOM-DARE Workshop (San Jose, Costa Rica, 17-28 July 2000) and the International Data Rescue Meeting (Geneva, 11-13 September 2001). WCDMP No. 49.