Data Rescue in the Southeast Asia and South Pacific Region

Challenges and Opportunities

BY CHER M. PAGE, NEVILLE NICHOLLS, NEIL PLUMMER, BLAIR TREWIN, MIKE MANTON, LISA ALEXANDER, LYNDA E. CHAMBERS, YOUNGEUN CHOI, DEAN A. COLLINS, ASHMITA GOSAI, PAUL DELLA-MARTA, MALCOLM R. HAYLOCK, KASIS INAPE, VICTOIRE LAURENT, LUC MAITREPIERRE, ERWIN E.P. MAKMUR, HIROSHI NAKAMIGAWA, NONGNAT OUPRASITWONG, SIMON MCGREE, JANITA PAHALAD, M.J. SALINGER, LOURDES TIBIG, TRONG D. TRAN, KALIAPAN VEDIAPAN, AND PANMAO ZHAI

Climate data, and their associated metadata, provide the fundamental building blocks for climate research and the development of climate products, applications, and services. In the past decade or so, the requirements of climate researchers to analyze and detect climate change and develop seasonal-to-interannual prediction systems have increased the importance of climate data. Climate data are also required for the preparation of climate models, which are widely used for verification of seasonal-to-interannual forecasts and to generate future climate scenarios. Maximizing the availability of digitized historical data and metadata is important for long-term climate monitoring, particularly for analyzing trends in the occurrence of extreme events where data quality becomes an even greater consideration.

Incorporating digitized climate data and metadata into regional databases suitable for enhanced quality control and analysis can benefit both regional and global climate studies and substantially improve our understanding of climate change. This is especially the case for heavy rainfall events and high daily maximum temperatures. National and regional analyses of extreme events have generally been restricted to the last 40–50 years because of a lack of earlier digital data, particularly on the crucial daily time scale. Even in the recent period, the lack of high-quality digitized daily data restricts attempts to monitor changes in climate extremes.

Metadata have been considered increasingly important in the development of homogeneous climate time series. Statistical methods alone are often inadequate in identifying and attributing inhomogeneities, and techniques are substantially improved through utilizing available metadata. The archiving of metadata in the country of origin, in both original and digital forms, is considered part of "best practice" for stations in the Global Climate Observing System surface network. While some countries have made progress toward the development of digital datasets of metadata (e.g., Australia and the United States), and others have published detailed station histories for key locations, metadata archives are often inadequate, even in developed countries.

The WMO's Data Rescue project, a component of the World Climate Data and Monitoring Programme

AFFILIATIONS: PAGE—CSIRO, Aspendale, Australia; NICHOLLS, MANTON, AND CHAMBERS—Bureau of Meteorology Research Centre, Melbourne, Australia; PLUMMER, TREWIN, DELLA-MARTA, COLLINS, AND PAHALAD—National Climate Centre, Bureau of Meteorology, Melbourne, Australia; ALEXANDER—Hadley Centre, Exeter, United Kingdom; CHOI—Korean Meteorological Administration, Seoul, Republic of Korea; GOSAI AND SALINGER—National Institute of Water and Atmospheric Research, Auckland, New Zealand; HAYLOCK—Climatic Research Unit, University of East Anglia, Norwich, United Kingdom; INAPE—NWS, Konedobu, Papua New Guinea; LAURENT—Météo-France, Tahiti, French Polynesia; MAITREPIERRE—Météo-France, Noumea, New Caledonia; MAKMUR—Meteorological and Geophysical Agency, Jakarta, Indonesia; MCGREE—Fiji Meteorological Service, Nadi, Fiji; NAKAMIGAWA—Japan Meteorological Agency, Climate Prediction Division, Tokyo, Japan; OUPRASITWONG—Meteorological Department, Bangkok, Thailand; TIBIG—Philippine Atmospheric, Geophysical, and Astronomical Services Administration, Quezon City, Philippines; TRAN—Climate Research Center, Hanoi, Vietnam; VEDIAPAN—Malaysian Meteorological Service, Petaling Jaya, Malaysia; ZHAI—National Climate Centre, Beijing, China
CORRESPONDING AUTHOR: Cher Page, CSIRO Atmospheric Research, Private Bag 1, Aspendale, Victoria, Australia, 3195
E-mail: Cher.page@csiro.au
DOI: 10.1175/BAMS-85-10-xxxx

(WCDMP), helps countries manage and preserve climate data and transfer observations from paper records to digital form. The WMO (in WCDMP Report No. 49) defines data rescue as "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," and adds that

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

Data-rescue projects have been undertaken in several parts of the world. For example, in a project funded primarily by Belgium, millions of documents from about 50 African countries were saved on microfiche and microfilm. Following the recommendations of the International Data Rescue meeting in Geneva in September 2001, Vietnam became the first country to receive, through the WMO Voluntary Cooperation Programme, equipment and training to start creating a digital-images archive of its climatological paper documents using optical scanners and digital cameras. Data will also be key-entered into a climate database.

The January 2001 WGCCD regional workshop on climate extremes in the Caribbean demonstrated what can be achieved through regional collaboration and relatively limited resources. Following the Caribbean workshop, several countries increased efforts to digitize their available paper archives, and the workshop, concluded Peterson et al. (in the *Journal of Geophysical Research*) "clearly fostered considerable interest and enthusiasm for daily data and data archaeology."

This paper will explore the opportunities and challenges of identifying and "rescuing" climate observations in Southeast Asia and the South Pacific. We believe the situation in this region, and the lessons to be learned, mirror those in many other regions.

ABOUT THE APN WORKSHOP SERIES

The first three Asia Pacific Network for Global Change Research (APN) Workshops on Indices and Indicators for Monitoring Trends in Climate Extremes (in 1998, 1999, and 2001) provided the Joint World Meteorological Organization (WMO) CCI/CLIVAR Working Group on Climate Change Detection (WGCCD) with a model for regional climate-change extremes workshops around the globe.

With a continued focus on hands-on development of climate extremes indices for many countries in the region, the fourth workshop, held 11–13 December 2002 at the Bureau of Meteorology in Melbourne, Australia, had many similarities to the previous three but included a greater focus on the precious resource upstream from the indices—climate data. Greater emphasis was placed on assessing the availability of data in the region, including historical meteorological station metadata (i.e., information such as changes in station location, surrounding environment, instrumentation, etc). The discussions led to this essay.

CURRENT STATUS OF CLIMATE RECORDS IN THE REGION. Each country in Southeast Asia and the South Pacific has climate data in paper and digital format. The amount of data available in each format differs among countries.

Precipitation, maximum and minimum temperature, and other climate variables have been recorded daily in most countries. In addition, some data are available at 3-hourly intervals for a number of stations. Monthly data (e.g., means, totals, and extremes within a month) are composites of daily and hourly records. In some cases, the original daily data are not available, but monthly data have been located and digitized.

Precipitation is the most commonly recorded climate variable, and during the twentieth century the number of precipitation stations in the region has increased (Fig. 1), while the amount of digitized records for temperature (Fig. 2) probably represents an upper limit on the volume of digitized records for other climate variables in most countries. French Polynesia, Korea, Indonesia, New Caledonia, New Zealand, and Thailand have digitized most or all daily precipitation and daily temperature data available on paper within their respective countries. China and the Philippines have digitized records from 1951 and Japan from 1961. Data are available for an additional 1300 automatic weather stations in Japan that are not represented in Fig. 1 or Fig. 2. The Japanese Meteorological Agency is currently digitizing the climate data of meteorological offices prior to 1961. For one period in Australia, digital records exceed the number of paper records. This is due, in part, to the use of automatic weather stations from which digital data are directly transferred into a database; thus, no paper records are produced.

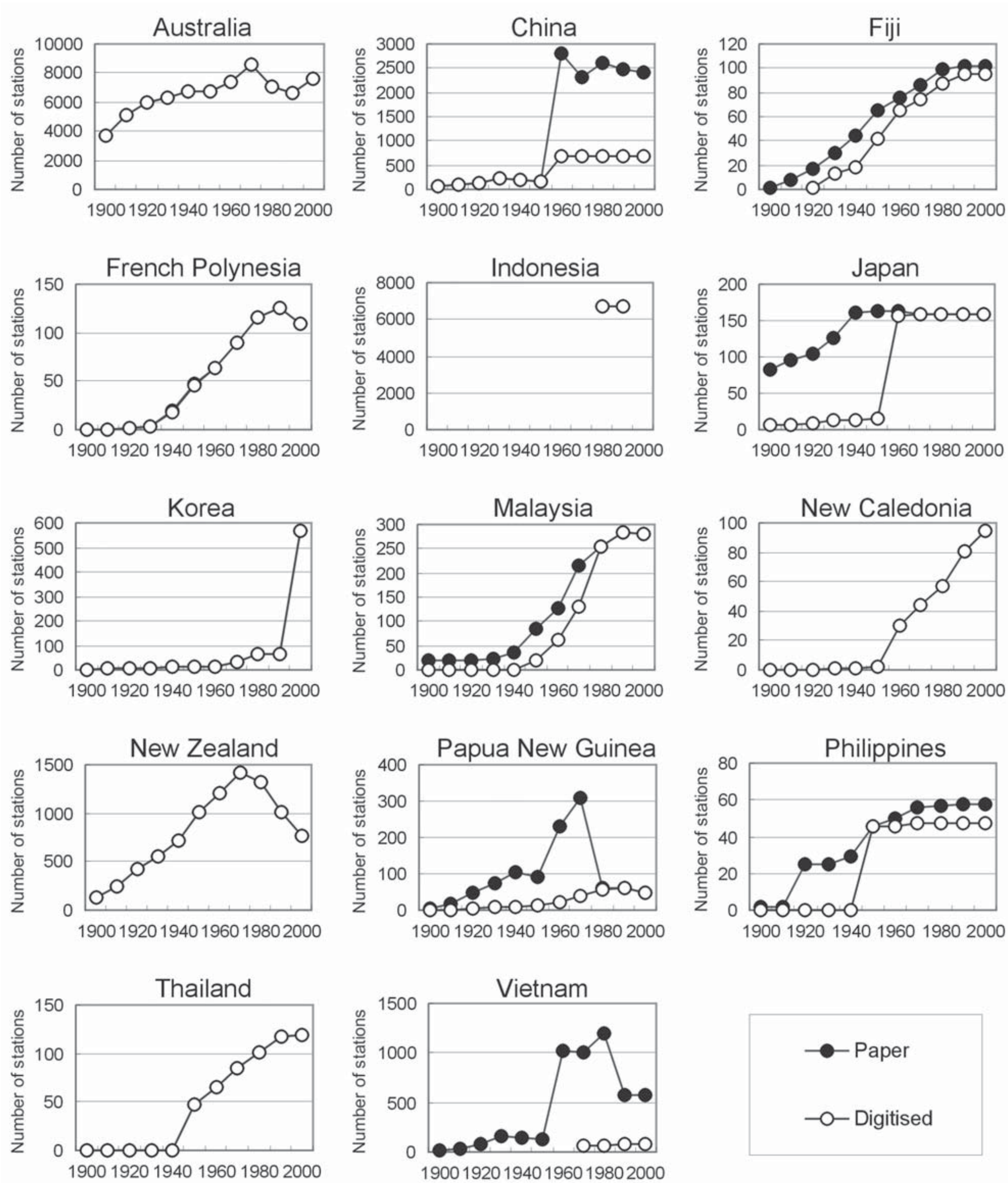

Fig. 1. For each country, the number of stations that have historical daily precipitation data available in paper format (filled circles) and in digital format (open circles), as sampled for the first day of every decade, from 1 Jan 1900 through 1 Jan 2000, and for each country. Increases or decreases in the number of stations during the following 10 yr to 1 Jan of the following decade have not been counted.

In the remaining countries, climate data exist in paper format only for a number of stations, and urgent effort is needed to ensure the digitization of these data.

OPPORTUNITIES TO RESCUE DATA. Demonstrating the benefit of digitized data to the user communities can help the effort to preserve data. For example, the Computerising the Australian Climate Archives (CLIMARC) project, which computerized historical daily and hourly climate data for 51 stations, was recently completed by the Bureau of Meteorology in Australia. The project was funded by agricultural research and development groups, which recognized that extended digital records would be useful in agricultural simulation models, seasonal-to-interannual prediction, and detection of changes in climate extremes. Two preliminary steps will be necessary prior to realizing the opportunities in Southeast Asia and the South Pacific:

1) *Conduct a systematic search for data and metadata.* In some countries in the region, one or more organizations take observations. These organizations will need to work together in developing databases of all station data and metadata in each country. For countries with colonial histories, another source of possible data will be the former colonizing country. Between them, Australia and New Zealand have historical data for Papua New Guinea, the Solomon Islands, Kiribati, Tuvalu, Tokelau, Samoa, Tonga, Niue, and the Cook Islands. The United Kingdom has some data for China, Fiji, and Singapore. As a starting point, countries will need to compile inventories of their records held both within the country and elsewhere. "Roving missions" by a few data experts may assist this process.
2) *Establish guidelines and systems to support data rescue efforts in the region.* Guidelines on data rescue must be consistent with an organization's broader data and information policy. Support for regional efforts to preserve data is urgently needed and this requires convincing funding bodies of the importance of data rescue. Following the inventory of available data, countries in the region will need people to transfer observations from paper records to computer databases and/or to microfilm or electronic image form. Compared to the ensuing benefits, only a relatively small investment in computing infrastructure and technical support is required.

CHALLENGES TO DATA RESCUE. Representatives at the fourth APN workshop identified five challenges to establishing a continuous record for meteorological stations within the region. These were: decaying paper records, missing records, lack of funds and resources, maintenance of the current network, and obsolete technologies.

Decaying paper records. Climate observations have traditionally been recorded on paper and later transferred to a central or regional office. An air-conditioned building is required for optimum storage of these paper records. Currently, one-third of all countries in the region do not have appropriate storage for all their paper records. The hot and humid climate in many countries increases the rate of decay of the paper records. For example, in the Philippines, parts of many bound volumes turn to dust when opened, so a program involving the use of microfilm and a digital camera could be used to preserve available historical data. With relatively modest funding, the remaining paper records within Southeast Asia and the South Pacific could be secured.

Missing data. There are numerous reasons why gaps appear in the climate records. The data may not have been recorded, may have been lost or destroyed over time, or may be spread across several different organizations or countries. In this region, for example, meteorological stations were suspended or closed during times of war. Observations at some stations in Korea are missing from between 1950 and 1953 and in Vietnam from between 1945 and 1954. WMO reports (in WCDMP Report No. 49) that the Hydrometeorological Data Centre in Vietnam is storing paper records in more than 30 rooms. On occasion, data have been lost due to organizational restructuring or policy decisions. Natural disasters such as the floods associated with Cyclone Tracy in 1974, which destroyed some files in the Northern Territory of Australia, are a constant threat throughout the region. Even with modern station networks, gaps appear due to factors such as instrument/sensor breakdown or, in the case of automatic weather stations, communication outages.

Lack of funds. For many countries, a lack of resources is a considerable challenge to digitizing historical observations and maintaining observational networks and related infrastructure. In many cases, digitization of records has occurred on an ad hoc basis or when paid for by an external client.

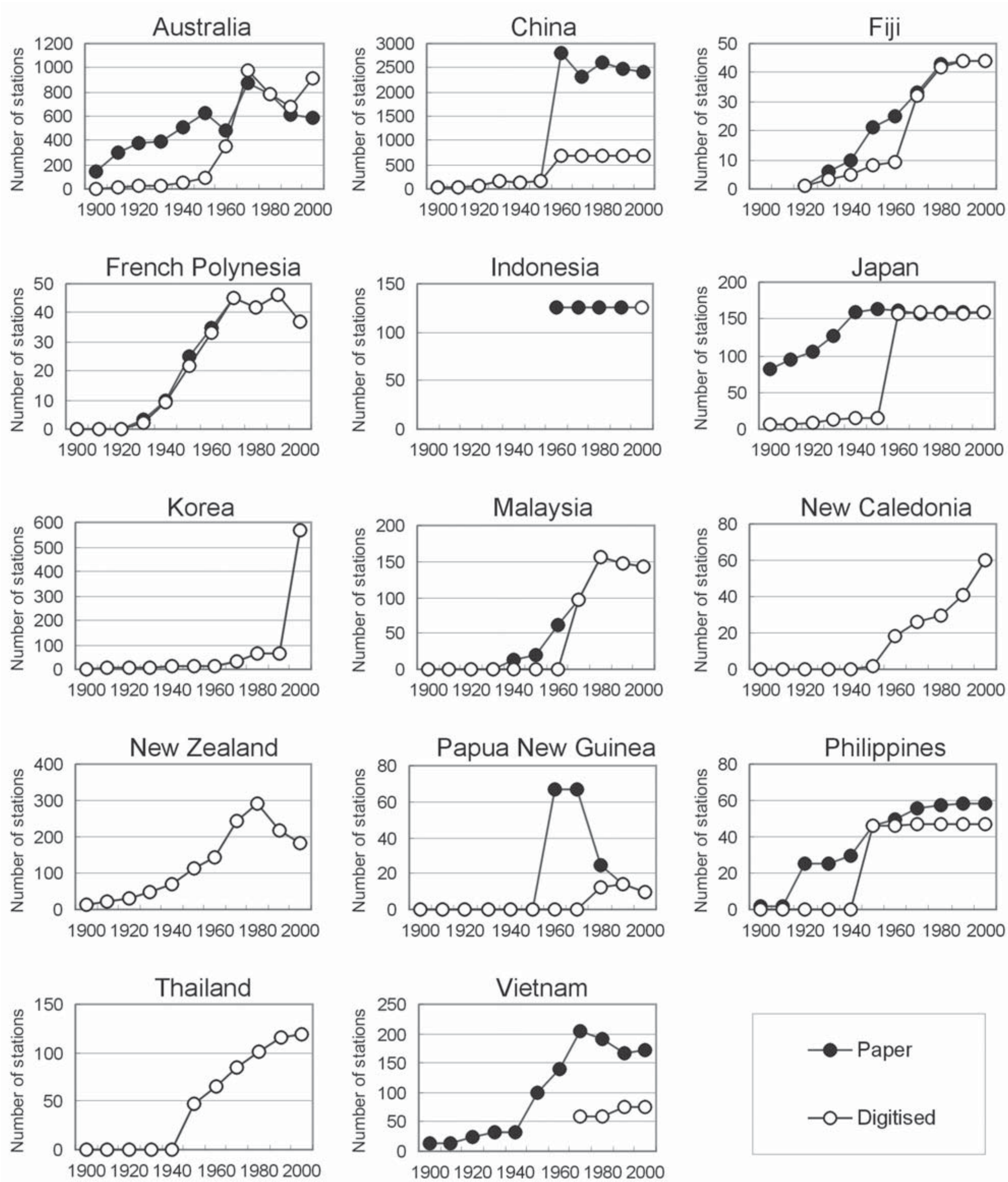

Fig. 2. Same accounting method as in Fig. 1, but for historical daily temperature data.

Maintenance of network. Maintaining the observational network is a challenge for many countries in Southeast Asia and the South Pacific. The number of cooperative and volunteer observers is declining and, in some countries, those who currently volunteer are increasingly requesting payment for their services. Much work remains before many automatic weather station (AWS) networks fully meet the climatological standards discussed by Plummer et al. at the Third International Conference on Experiences with AWS's in Malaga, Spain, in 2003. Participants at the APN workshop reported mixed experiences with AWS's. In 2000, South Korea moved its entire observing network from manual observations to AWS's. Papua New Guinea has two AWS's, neither of which is currently operational. Due to the remoteness of stations, particularly in Fiji and French Polynesia, there are often very long delays in repairing AWS's when they malfunction. Indonesia collects the AWS data via subcollection points using radio technology and recently had some trouble with this hardware. Of course, problems of maintaining a network suitable to the needs and standards of climatology (e.g., ensuring data continuity and homogeneity) are not exclusive to Southeast Asia and the South Pacific, but will likely remain a challenge to all countries.

Obsolete technologies. There are potential problems even for data that have found their way into computer-compatible form. Computer media eventually become obsolete and incompatible with modern archiving and processing facilities. Several factors are important in this regard, including the choice of media and the assurance that forward data migration strategies are included in information management plans. Support of software and databases to store and analyze climate data needs to be maintained throughout the region.

CONCLUSIONS. There is still much to learn about the amount of climate data available in Southeast Asia and the South Pacific. Nevertheless, through collecting information on available data sources and exploring some of the opportunities and challenges that await, this paper has intended to provide a basis from which efforts to improve data availability and accessibility can be made.

Through the collective efforts of national meteorological and hydrological services (NMHS's), the WMO Commission for Climatology, the APN, Pacific Island Global Climate Observing System, and other bodies interested in furthering the understanding of regional and global climate change, it is hoped that progress can be made in this regard.

As we have documented here, historical climate records in many (perhaps most) countries are undigitized and only available in paper form. Many of these data are stored inappropriately or are now located in organizations outside of NMHS's—sometimes in other countries. Metadata are generally poorly managed and are probably more vulnerable to damage from poor storage or meteorological (and nonmeteorological) disasters than the conventional climate data, yet they are essential if we are to adequately monitor a changing climate. The ability to adjust data for variations in instrumentation and exposure requires metadata. Many countries—even developed countries—do have some archives of metadata, but these are almost invariably stored improperly and difficult to access.

Compared to the operating and asset costs of many other activities undertaken by NMHS's, the costs of preserving climate data and metadata are relatively small and, given the technology now available, the task is not especially complex. The provision of funding for just one person in each country—plus a high-quality personal computer with scanner, printer, and software—to focus on preserving climate records would make substantial inroads into the problem. However, such efforts have not yet been given adequate attention in Southeast Asia and the South Pacific. Unless there is a change in priorities and recognition of the importance of this activity, it seems likely that much data (and even more metadata) will soon be lost forever, and not only in developing countries. Such a loss will mean that our ability to document and monitor climate variability and change, especially with regard to climate extremes, will be eternally compromised.

ACKNOWLEDGMENTS. This work arose from a workshop held in Melbourne, Australia, on 11–13 December 2002 with kind support from the Asia Pacific Network for Global Change Research. Rod Hutchinson and Robin Hicks (National Climate Centre, Bureau of Meteorology, Australia) provided Australian data for Figs. 1 and 2.

FOR FURTHER READING

Arnfield, J. D., 2001: A flexible system to manage and query NOAA station history information. *17th Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology*, Albuquerque, NM, Amer. Meteor. Soc., 468–472.

Basher, R., B. Collen, B. Fitzharris, J. Hay, B. Mullan, and J. Salinger, 1992: *Preliminary Studies for South Pacific Climate Change: Present Climate Change and Its Impacts, Data Resources and Scenario Possibilities.* New Zealand Meteorological Service, 70 pp.

Boodhoo, Y., cited 2003: Second circular letter of the president of the WMO Commission for Climatology.

[Available online at www.wmo.ch/web/wcp/ccl/presclets/prescl_jan03_en.pdf.]

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, Australia, Bureau of Meteorology, 49.

Collen, B., 1992: *South Pacific Historical Climate Network. Climate Station Histories: Part 1: Southwest Pacific Region*. New Zealand Meteorological Service, 114 pp.

Daan, H., 2002: Guide to the GCOS surface and upper-air networks: GSN and GUAN. WMO/TD No. 1106. World Meteorological Organization, 36 pp.

Lane, J., 1995: Sites Database: Functional requirements. Australian Bureau of Meteorology Internal Rep., 16 pp.

Manton, M. J., and Coauthors, 2001: Trends in extreme daily rainfall and temperature in Southeast Asia and the South Pacific: 1961–1998. *Int. J. Climatol.,* **21**, 269–284.

Nicholls, N., 1995: Long-term climate monitoring and extreme events. *Climatic Change,* **31**, 231–245.

Peterson, T.C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.,* **18**, 1493–1517.

——, and Coauthors, 2002: Recent changes in climate extremes in the Caribbean region. *J. Geophys. Res.,* **107** (D21), 4601.

Plummer, N., D. Collins, K. Wong, and P. Della-Marta, 2000: Review of current status and needs of the GCOS surface network and the GCOS upper air network. *Report of the Pacific Islands Regional Implementation Workshop on Improving Global Climate Observing Systems,* WMO/TD No. 1038, World Meteorological Organization, 47–54.

——, and Coauthors, 2003: Progress of automatic weather stations in meeting the needs of climate. *Third International Conf. on Experiences with AWSes*, Malagas, Spain, Spanish National Meteorological Service, CD-ROM.

World Meteorological Organization, 2002: Report of the CLICOM-DARE Workshop (San Jose, Costa Rica, 17–28 July 2000) and the Report of International Data Rescue Meeting (Geneva, 11–13 September 2001). WCDMP Rep. No. 49. WMO, 7 pp.