J8.10 ANALYSIS OF DATA EXCHANGE PROBLEMS IN GLOBAL ATMOSPHERIC AND HYDROLOGICAL NETWORKS

Stephan Bojinski and Alan R. Thomas
GCOS Secretariat c/o World Meteorological Organization
7bis, Avenue de la Paix; 1211 Geneva 2; Switzerland

1. EXECUTIVE SUMMARY

The United Nations Framework Convention on Climate Change (UNFCCC), through the Conference of the Parties (COP) and its Subsidiary Body for Scientific and Technological Advice (SBSTA), has repeatedly identified the urgent need for improving the availability and access to systematic global observations of climate variables for the purposes of the Convention. The Global Climate Observing System (GCOS) Secretariat is submitting this report to the twentieth session of SBSTA (June 2004) as a preliminary analysis of availability and accessibility of atmospheric and hydrological data relevant to climate, from the perspective of the monitoring and archiving centres. The analysis of the broader issue of problems in data exchange and potential remedies, which SBSTA[1] requested at its 18th session, will require further consultation with the World Meteorological Organization and others, such as the *ad hoc* Group on Earth Observations (GEO)[2]. This analysis is based on an empirical survey carried out among managers of data archive and monitoring centres, primarily those associated with the GCOS baseline networks.

Despite a number of deficiencies, the infrastructure for global data exchange exists for meteorological, including atmospheric composition, climate variables. Designated monitoring and data archive centres are in place that allow easy access to climate datasets. In the hydrological domain, a designated international data centre is in place only for river flow properties, and large gaps exist in the receipt of data from existing networks. Substantial effort is still required in terms of hydrological data standardization (e.g., improved metadata), standard operating practices, data policy, infrastructure, and awareness. The situation is better in the related cryospheric domain, where designated monitoring and archive centres are operating.

The major problems and challenges identified in this study include:

- Reluctance of some countries to exchange data, due to a number of reasons;
- Need to improve the awareness of climate requirements for global data exchange;
- Technical problems in preparation, transmission and receipt of climate messages;
- Resource limitations in developing countries and, to some extent, in archive and monitoring centres;
- Data and metadata standardization and data stewardship.

The existing situation could be improved by increasing awareness, on the political and scientific level within countries and funding agencies, of the importance and benefits of the free and unrestricted data exchange for climate purposes. Continuing encouragement is required from international bodies (e.g., WMO, UNFCCC) for efforts by the international data centres to obtain permission from countries for the release of data and the rescue of historical climate records. A clearer expression of requirements for global climate data by the climate science community would be useful in discussing with countries the needs to exchange data.

Data management, stewardship, and long-term access to data were seen as crucial by the data centres. The international data archive centres are major assets in providing for free access and distribution of climate data and products, and therefore require sustained long-term funding. In addition, the monitoring centres are essential to maintain the quality and regular receipt of data, through the capability to provide timely and appropriate feedback to observers. Furthermore, there was an emphasis on the need for consistent time series of climate observations, which are made more feasible by technological progress.

[1] Conclusion # 7 from the 18th session of SBSTA was: "To better understand the barriers to improving the receipt, at global data centres, of data from atmospheric and hydrological networks, the SBSTA invited the GCOS Secretariat to prepare, in consultation with the World Meteorological Organization (WMO), an analysis of specific problems and of options to remedy them, for consideration by the SBSTA at its twentieth session. The SBSTA further invited the GCOS secretariat to comment, in its report, on the accessibility of data from global data centres."

[2] The Terms of Reference for the *ad hoc* Group on Earth Observations (GEO) state that *inter alia*, "The GEO will seek in its work to …exchange observations recorded from *in situ*, aircraft, and satellite networks, dedicated to the purposes of this Declaration, in a full and open manner with minimum time delay and minimum cost, recognizing relevant international instruments and national policies and legislation."

2. BACKGROUND

The UNFCCC has repeatedly identified the need for improving the availability and access to systematic global observations of climate variables as an urgent requirement of the Convention. Improved data is needed for understanding the anthropogenic causes of climate change, identifying concentrations, sources and sinks of greenhouse gases, assessing climate variability and extreme events, and tracking climate changes over time. All require a sound basis of long-term observational climate data of high quality. Major deficiencies have been singled out by the Intergovernmental Panel on Climate Change (IPCC, 2001) and the Global Climate Observing System (GCOS, 2003), and can be summarized in two major categories: (1) inadequacies in the global observing systems for climate, and (2) unavailability of the observational data that have been collected due to problems in the exchange, management, stewardship, and access to the data. For example, IPCC (2001) states in general terms that over one third of the Earth's land mass is not covered by historical climate observations due to a lack of data exchange.

This study concentrates on the latter category by highlighting specific problems in the exchange of data in the atmospheric and hydrological domains for the global *in situ* networks, based on an empirical survey of managers of data archive and monitoring centres. This work is in response to an invitation to the Global Climate Observing System (GCOS) Secretariat from the UNFCCC's Subsidiary Body for Scientific and Technological Advice (SBSTA) at its 18th session in June 2003, in order "to better understand the barriers to improving the receipt, at global data centres, of data from atmospheric and hydrological networks", to "prepare, in consultation with the WMO, an analysis of specific problems and of options to remedy them", and to "comment, in its report, on the accessibility of data from global data centres".

The primary emphasis of the study was to survey the data archive and monitoring centres associated with the GCOS baseline networks, which routinely measure:

- Meteorological climate variables,
- Atmospheric composition variables (not addressed in this summary report), and
- Hydrological and related variables,

as a part of a global network, mainly for the purposes of the climate community. This includes real or near-real-time observations as well as the rescue and receipt of historical data records, data monitoring, systematic storage, and appropriate interfaces for data access. The monitoring and quality control aspect is especially important for climate research and assessment, where requirements on data quality are high for issues such as trend detection or initial conditions for climate models.

In light of the requirement to address the data exchange and access issues highlighted in the Implementation Plan for Global Observing Systems for Climate currently being developed by GCOS, as well as in the Implementation Plan of the Global Earth Observation System of Systems (GEOSS), the insights given by this study can be used as specific examples for the case of atmospheric and hydrological climate data (in support of the UNFCCC) in the GCOS context.

3. SCOPE OF STUDY

This study focuses primarily on the global data archive and monitoring centres associated with the GCOS baseline networks, dealing with *in situ* atmospheric Essential Climate Variables (ECVs), and, where available, *in situ* hydrological ECVs.

Managers of the respective monitoring and data archive centres were contacted to assess specific problems in the receipt of near and non real-time data, e.g., due to incorrect data formatting, poor measuring equipment, or internal resource constraints. Monitoring and quality control of data streams were examined in terms of feedback mechanisms to ensure regular high-quality data production. Issues around the accessibility of obtained and monitored data, by internet or other means, were also analysed.

The accessibility issue has been assessed from the data providers' perspective, leaving aside the view of scientists or other users.
Figure 1 indicates the scope of this study within the overall data flow of the GCOS baseline networks.

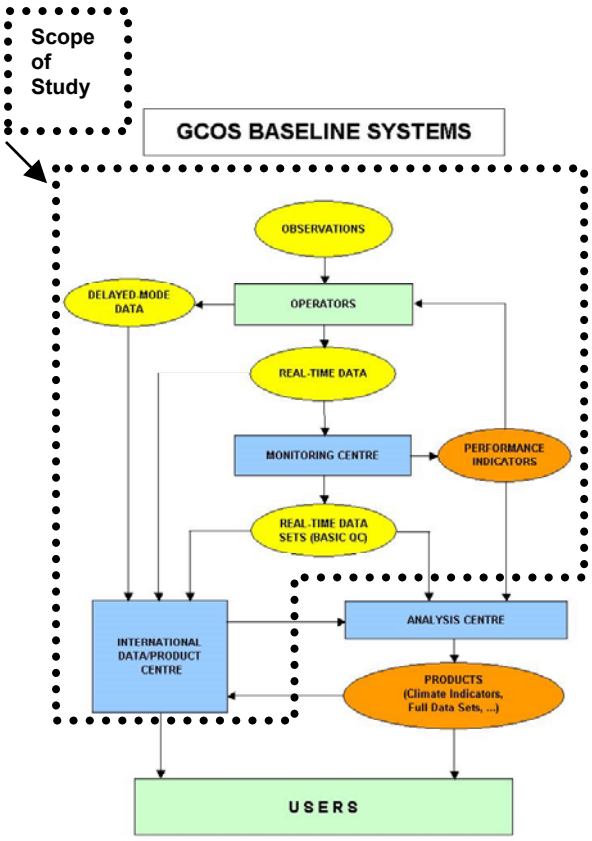

Figure 1: Schematic illustration of the scope of this study, based on the example of GCOS baseline networks (GCOS, 2003).

4. DISCUSSION

4.1 Data Exchange in the GCOS Surface Network

Essential Climate Variables: Surface air temperature, Precipitation, Surface air pressure

Archive/Lead Centre - US National Climatic Data Center / World Data Center for Meteorology

The US National Climatic Data Center (NCDC) is the archive centre for the GCOS Surface Network (GSN) and GCOS Upper-Air Network (GUAN), as well as being a WMO/CBS lead centre for GSN and GUAN data. Its responsibilities include contacting countries with the purpose of obtaining historical datasets, digitizing paper records, reformatting received data, overseeing GSN/GUAN monitoring activities (see sections below), and making all received data available.

NCDC undertakes activities to rescue and digitize historical climate records, in order to extend available time series for surface air temperature and precipitation as a part of its Global Historical Climatology Network (GHCN) dataset. This has involved digitizing station history files, old books, maps and paper records.

Currently, some historical GSN data for 407 out of the 981 GSN stations worldwide have been made available in response to requests from WMO. Historical data for another 364 GSN stations are available to NCDC from other archival sources.

Figure **2** displays the total number of GSN stations in comparison with the total amount of GSN stations for which data are actually available from the GHCN dataset, for mean surface air temperature. Curves for precipitation are similar in shape. The total number of stations belonging to the GSN increased steadily until 1960, when the growth in number reached a plateau near today's 981 stations. Monthly and daily data holdings increased similarly, with monthly data generally outnumbering daily data until the past decade. The dip in monthly data around 1990 is due to delays in updating the GHCN, which depend on retroactive data compilations, such as the World Weather Records, which are processed in decadal steps (the last one 1991-2000). The decrease in monthly data around 1970 is mainly due to the suspension of extensive efforts to digitize historical data at that time. The recovery of monthly data in the mid-1990s can be attributed to facilitated data exchange, following the WMO initiative of CLIMAT data transmission over the GTS.

Data can be ordered through NCDC's website[3]. A new version of the website is due to be completed by September 2004.

For stations sending daily (SYNOP) data to the data archive centre, but not the desired monthly (CLIMAT) data, activities are underway within the WMO World Climate Programme (WCP) and World Weather Watch (WWW), in coordination with GCOS, to improve this situation. These include the detailed checking of errors and automating the process of preparing CLIMAT messages.

Major problems and challenges are as follows, including some suggested remedy options:

(1) Lack of resources at data archiving centre to rescue and archive climate data

Many historical climate records are not widely available, due to a lack of demand, interest, or human and technical resources. Substantial effort is required to contact countries for historical data records and convince them to release their data. Progress is being made, but slowly. Explicit requests have to be issued to countries by WMO

[3] See list of Internet References.

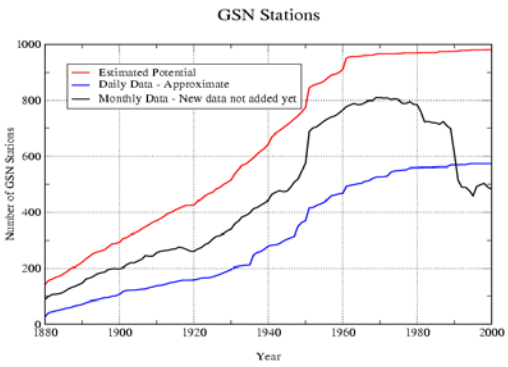

Figure 2: Historical daily and monthly mean surface air temperature data for GSN stations available from the NCDC GHCN dataset. 'Estimated Potential' (upper line) is the total number of GSN stations existing historically, while the monthly data (middle line) and daily data (lower line) are stations for which data actually available from NCDC. The drop in monthly data in the 1990s is mainly caused by delays in updates of the GHCN data archive.

and its lead centres (e.g., NCDC), for the release of climatological data, which are otherwise not distributed freely. National points of contact are being identified, but not all countries have done so. More resources for establishing and maintaining contact with countries are urgently needed. In addition, resources are required to harmonize and organize daily and monthly GSN data available from different sources and in different formats into the data archives.

(2) Restriction of data access to special interest groups

Many countries do not regularly submit climate data to NCDC. In some cases, bilateral agreements between national organizations and the NCDC have proved quite successful in getting data released. Progress is slow however. Obtaining data over a region in many cases requires great effort to contact countries individually, and data release is sometimes restricted to specific purposes. Currently, a joint WMO Commission for Climatology/CLIVAR Expert Team for Climate Change Detection, Monitoring and Indices, in cooperation with IPCC and GCOS, is working with scientists to develop a set of indices for analyses of climate extremes. These indices will be calculated using national data, analyzed on a regional basis and then synthesized into a true global analysis for use in the next IPCC Assessment.

(3) Lack of human and technical resources in countries, or at stations, to rescue data through digitizing paper records and quality-controlling and archiving the data

Monitoring Centres – Japan Meteorological Agency (JMA) and Deutscher Wetterdienst (DWD)

The GSN monitoring function is being shared by JMA (monitoring of temperature data) and DWD (monitoring of precipitation data). JMA is also a WMO/CBS lead centre for GSN data. The monitoring centres check the timeliness and completeness of the submitted datasets and identify non-reporting stations. A well-defined consistency and quality control process, including manual interaction and automated steps, ensures high quality of data and allows regular reporting of network performance on the internet. The reception rate of CLIMAT monthly reports has increased from 55% in early 2001 to 65% in late 2002, and to around 68% at the end of 2003.

Major problems and challenges are as follows, including some suggested remedy options:

(1) Low GSN performance in some regions

Low reception rates of GSN stations have consistently been noted in Africa (WMO Region I) and Antarctica.

(2) Incorrect data format

Regular data receipt is often subject to erroneous coding of the CLIMAT reports as well as the station metadata.

4.2 Data Exchange in the GCOS Upper-Air Network

Essential Climate Variables: Upper-air temperature, Air pressure, Wind speed and direction, Water vapour

Archive/Lead Centre – US National Climatic Data Center / World Data Center for Meteorology

The US National Climatic Data Center (NCDC) is the archiving and lead center for the GCOS Upper-Air Network (GUAN). Data receipt and accessibility conditions are by and large comparable to those described in the previous section for GSN data. Attempts to rescue historical data records and make them available by NCDC are a permanent priority.

Major problems and challenges are as follows, including remedy options where proposed:

(1) National reluctance to release data and lack of awareness of the need for historical data for climate purposes (e.g., IPCC assessments, regional impact studies)

The refusal of certain countries to make the data collected under their auspices available to the

global data centre (except possibly for the daily global weather forecast) poses a serious challenge for climate studies. This is partly due to unclear national responsibilities and commercialization issues. For example, an effort led by the Meteorological Service of France (Météo France) rescued, and digitized, daily synoptic upper-air (and surface) data measured in 14 West African countries ("African database"). The data have not been released in their entirety; to date only four countries agreed to allow free access to these data.

Monitoring Centres (UK Met Office Hadley Centre, ECMWF)

The Hadley Centre monitors the monthly data for the GCOS Upper-Air Network (GUAN). Additionally, daily GUAN data are registered and checked by the European Centre for Medium-Range Weather Forecasts (ECMWF). Automated quality checks are used to determine the need for manual quality checks. Overall, a small improvement in the reporting of CLIMAT TEMP reports has been indicated since 1998. Timely documentation of monthly monitoring results and procedures are made available through a dedicated website.

Major problems and challenges are as follows, including some suggested remedy options:

(1) Problems in receiving monthly GUAN data

Poor coding and missing data are attributed by the monitoring centre to telecommunications failure. The monitoring centre is unable to quantify whether a station is taking data, but not reporting through the telecommunications system, or is simply silent.

(2) Significant gaps in data coverage

Many stations in Africa, South America and in small-island regions are not operating routinely due to the high cost of operations, especially consumables.

(3) Lack of funding for station operation – consumables

In general, regular operation of a GUAN station requires sustained funding for the purchase of relatively expensive balloons and rawinsondes (around USD 200 for each launch).

4.3 Data Exchange in River Runoff Networks

The Global Runoff Data Centre (GRDC) at the German Federal Institute of Hydrology (BfG), Koblenz, is the only identified international data centre dealing with river runoff data. The Centre collects and disseminates river discharge data from 3800 stations, monitoring 2900 rivers on a global scale. The national data providers (e.g., hydrological services) generally carry out the data quality monitoring, while GRDC, after performing plausibility checks, makes the data available electronically upon request within 8 working days. GCOS, through its Terrestrial Observation Panel for Climate (TOPC), has recently defined a Global Terrestrial Network for Rivers, which will observe river discharge into the ocean. Figure 3 shows the amount of data available from river stations, mostly available via the GRDC website. Regular inflow of station data is shown with a prominent peak of available data around 1980, and far less total available station data for the late 1990s. GRDC operations statistics show that it currently makes around 240 contacts per year in 40 countries aimed at acquiring missing data, with one third resulting in successful data acquisition.

Major problems and challenges are as follows, including some suggested remedy options:

(1) No institutionalized global exchange of river data

As only four countries worldwide send their data automatically, there is considerable room for improvement. National policies and lack of centralized institutional responsibility largely prevent data exchange, often supported by perceived or real commercial and strategic value of the data. Wider data exchange across boundaries is mostly restricted to the level of river basins or agreed upon in bilateral agreements. Received data at GRDC often show considerable time delays, e.g., when retrieved only from published hydrological yearbooks.

(2) No international standards for hydrology data

There are no established international standards on the acquisition of river data, the set of required metadata, data formats, and transmission modes. The definition of these standards is a high priority, recognized by the WMO and the UNFCCC, for example.

(3) Fragmented data holdings

Hydrological data in general, and river information in particular, are frequently stored in a distributed manner on national levels. Different institutions responsible for sectors like water, energy or agriculture, often keep their own data records without exchanging them at national levels. Common metadata standards are, even on national scales, the exception rather than the rule.

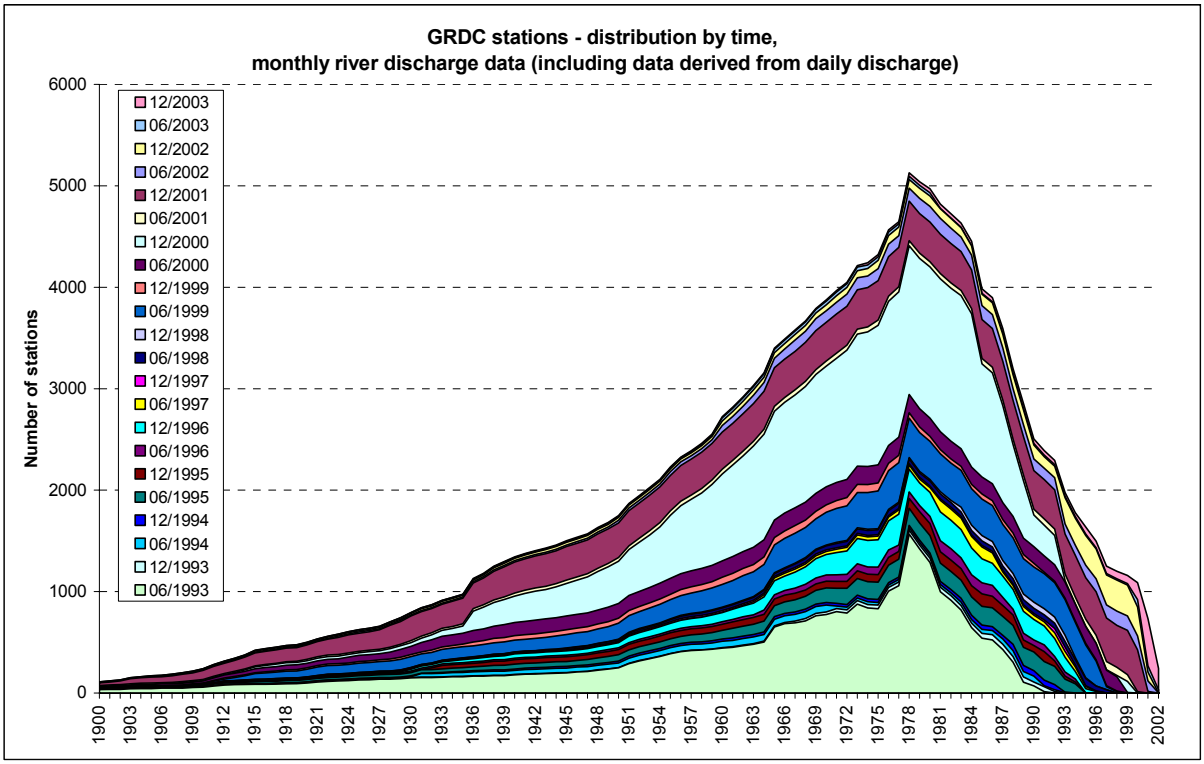

Figure 3: River discharge station data in GRDC between 1900 and 2002, showing incremental half-yearly increases from 1993 to 2003.

(4) Declining networks

A trend towards declining networks has been noted since the 1980s, due to instabilities in many places and economic constraints, occurring mostly in developing countries. Disparities exist between regions in terms of the density of observing river stations, especially in remote areas. For example, network cutbacks were particularly extensive in remote arctic areas, where up to 73% of river gauges were closed between 1986 and 1999 (Shiklomanov et al., 2002).

5. ACKNOWLEDGEMENTS

Acknowledgements are expressed to all contributors to this study who replied to the included survey, as well as the support by the GCOS Secretariat and the GCOS Steering Committee Chairman.

6. REFERENCES

GCOS, 2003: The Second Report on the Adequacy of the Global Observing Systems for Climate in Support of the UNFCCC. GCOS-82 (WMO/TD-No. 1143), 80pp.

IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA, 881pp.

Shiklomanov A.. *et al.*, 2002: Long-term Variability of the Pan-Arctic Hydrological Budget and the Decline in Hydrological Monitoring Networks in: ARCSS All-Hands Workshop, Bell Harbor International Conference Center, Seattle, USA, February 20, 2002.

7. INTERNET REFERENCES (24 May 2004)

GSN at NCDC – WDC Meteorology:
http://www.ncdc.noaa.gov/oa/climate/gsn/gsninfo.html
GSN Monitoring at JMA/DWD:
http://www.gsnmc.dwd.de
GUAN Monitoring at UK Met Office Hadley Centre:
http://www.guanweb.com
GRDC at BfG: http://www.bafg.de/grdc.htm

8. ACRONYMS

BfG	Bundesanstalt für Gewässerkunde (German Federal Institute of Hydrology)
CBS	Commission for Basic Systems
CLIMAT	Report of monthly means and totals from a World Weather Watch land station
CLIVAR	Climate Variability and Predictability
COP	Conference of the Parties (to UNFCCC)
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
GCOS	Global Climate Observing System
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GHCN	Global Historical Climatology Network
GRDC	Global Runoff Data Centre
GSN	GCOS Surface Network
GTS	Global Telecommunication System
GUAN	GCOS Upper-Air Network
ICSU	International Council for Science
IOC	Intergovernmental Oceanographic Commission (of UNESCO)
IPCC	Intergovernmental Panel on Climate Change
JMA	Japan Meteorological Agency
NCDC	National Climatic Data Center
SBSTA	Subsidiary Body for Scientific and Technological Advice
SYNOP	Report of surface observation from a World Weather Watch land station
TOPC	Terrestrial Observation Panel for Climate
UBA	Umweltbundesamt (German Federal Environmental Agency)
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNFCCC	United Nations Framework Convention on Climate Change
WCP	World Climate Programme
WDC	World Data Centre
WMO	World Meteorological Organization
WWW	World Weather Watch