Hydrological Data Rescue – the current state of affairs

MATTHEW FRY

Centre for Ecology and Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxon, OX10 8BB, UK

mfry@ceh.ac.uk

Abstract Hydrological data are costly to record and collect, in terms of both effort and resources. Historical hydrological data records are important both nationally and internationally for activities from water resources management to flood estimation and climate change modelling. The loss of data can have a significant impact on the ability to undertake these activities. Data rescue is the process of securing data at risk of loss through natural hazard, degradation or redundancy of storage medium, and providing access to data through digitisation and computerisation. A World Meteorological Organisation survey of National Hydrological Services worldwide requested information on current hydrological data rescue requirements. Results indicate huge volumes of data at risk. Many countries requiring data rescue are poorly represented within international flow archives. Data rescue efforts targeted towards these countries and towards capturing gauging station information would improve the data within these archives for a wide range of applications.

Key words data rescue, flow archives, international

INTRODUCTION

Historical hydrological data records are important at a national and international level: for process understanding, water resources management and modelling; climate change detection and climate modelling; flood modelling and prediction; and other hydrological and engineering activities. The loss of data can have a significant impact on the ability to undertake these activities and reduce the quality of results obtained from them. In addition, the loss of descriptive station information can mean that rating curves – sensitive relationships between water level and flow – and the quality of flow records, cannot be reassessed, and that the impact of changes to the catchment over the period of the record cannot be understood.

Hydrological data are costly, in terms of both effort and resources, to record and collect. Despite this, large volumes of data are lost due to inadequate archiving of collected data and also due to poor maintenance of data archives. A decline in the monitoring of the Earth's hydrology in the late 20th Century has been widely documented (e.g. World Bank, 1993; WMO, 1996; Giles, 2005). This has been seen in the neglect and abandonment of stations, reductions in budgets for field maintenance and inspection, and insufficient discharge measurements being made to adequately define rating curves. The result is a significant reduction in the coverage of the river monitoring network, and a reduction in the network of stations with long records (Vorösmarty, 2002).

In these circumstances there would appear to be a need to maximise the usage and impact of the hydrological data that has been captured in the past. There is also an increasing need for good quality hydrological data, and contextual metadata, within the international community, for the detection of potential climate change signals in rivers, and for the assessment, or calibration, of hydrological models to link to ever-improving Global Climate Models. But there is anecdotal evidence for an equivalent reduction in the budgets for, or efforts towards, the management of hydrological data archives and databases in some countries. It is vital to understand the level of hydrological data at risk, and to direct efforts towards safeguarding the most important data at the greatest risk.

DATA RESCUE INITIATIVES

To this end there have been several data rescue (often termed DARE) initiatives. Data rescue is the process of securing data at risk of loss through natural hazard, degradation or redundancy of storage medium, and providing access to data through digitisation and computerisation. Data rescue has been a major focus in meteorology for several decades, recently driven by the need for higher quality and further reaching 'reconstructions' of past climates for climate change detection, and climate modelling. The World Meteorological Organisation (WMO) started data

rescue in 1979 with a project, assisted by the Belgian Government, which successfully digitised over one million meteorological documents in northern and western Africa (WMO, 2002). The Atmospheric Circulation Reconstructions over the Earth (ACRE) project (www.met-acre.org) has focussed largely on the identification and digitisation of historic ship's logs in order to introduce valuable data from periods and locations of poor data coverage. These are two examples of large-scale data rescue projects; many data rescue initiatives are funded within individual National Meteorological Services, but provide resulting digitised data to the international community.

HYDROLOGICAL DATA RESCUE

The progress in meteorological data rescue has not been seen in the field of hydrology. FRIEND has not undertaken any specific hydrological data rescue activities, though it has collated regional databases for research purposes, e.g. European Water Archive, Southern Africa FRIEND river flow and spatial databases (Servat & Demuth, 2006). These databases constitute substantial archives of secured data, available for international research. The Global Runoff Data Centre (GRDC) has been hugely successful in accumulating river flow data, and achieving recognition from many countries of the need to share data. Many national and international projects have succeeded in providing capacity to manage hydrological data electronically. But the complexity of defining hydrological data quality is greater than for meteorological data, and the rescue and provision of river flow time series without data quality information is not sufficient to preserve the data. A precipitation or temperature measurement, with adequate validation against data from nearby locations, can generally be taken as accurate by a user of that data. But river flow data have numerous sources of potential error, for example instrumentation, measurements of cross-sectional area, and rating equations, and so can vary widely. An accuracy of 5 – 10 % can typically be achieved under good conditions (Hirsch & Costa, 2004), though is often not possible to obtain. Therefore an understanding of the factors that influence data quality is absolutely essential when utilising the data. In addition there are anthropogenic factors that can affect catchment runoff, such as reservoirs, increased abstraction or discharge. An understanding of

these factors, and how they have changed over the period of record, is also essential when using river flow data for almost any purpose.

However neither data quality information, nor data on human impacts, is generally available internationally, and is rarely stored alongside the river flow time series data at a national level. In addition there is a perception that the digitisation of national archives has only been undertaken piecemeal, with records from many, generally less operational, stations remaining in some earlier medium.

NATIONAL-LEVEL NEED FOR DATA RESCUE

However, as stated, current knowledge of both the level of data loss, and the need for more intensive hydrological data rescue, is based on anecdotal evidence. Between 2006 and 2008 the WMO Commission for Hydrology (CHy) surveyed its member National Hydrological Services (NHS) with the aim of producing a clearer picture of the extent of the hydrological data rescue problem. A series of questions was asked concerning the need for data rescue, the nature of data rescue required, and previous data rescue activities. The findings of the responses are summarised below.

Of 183 member countries there were 58 responses from NHS in 56 countries (30%). This was considered a good response rate for such a survey, and broadly indicative of the need globally. The nature of such surveys, and factors influencing a response, mean that countries requiring data rescue assistance could have been more active in responding than those with no need for data rescue. This paper aims to describe the need for hydrological data rescue through the responses given, but does not attempt to scale the numbers up to reflect total global demand. Table 1 shows the breakdown of responses by WMO Regional Association (RA), and the number of those countries specifying a need for data rescue.

Table 1 Data Rescue survey responses by WMO RegionalAssociation.

WMO Regional Association	RA 1 Africa	RA 2 Asia	RA 3 South America	RA 4 North America, Central America and the Caribbean	RA 5 South-West Pacific	RA 6 Europe	Total
Responses (countries)	16	4	5	6	10	15	56
Data rescue need	14	2	5	4	4	8	37
Countries requiring urgent data rescue	8	0	2	4	2	3	19

Data rescue need did not correspond to a country's income. Many developed countries, in Europe and North America, where hydrological data management systems are known to be well developed and resourced, identified data rescue needs.

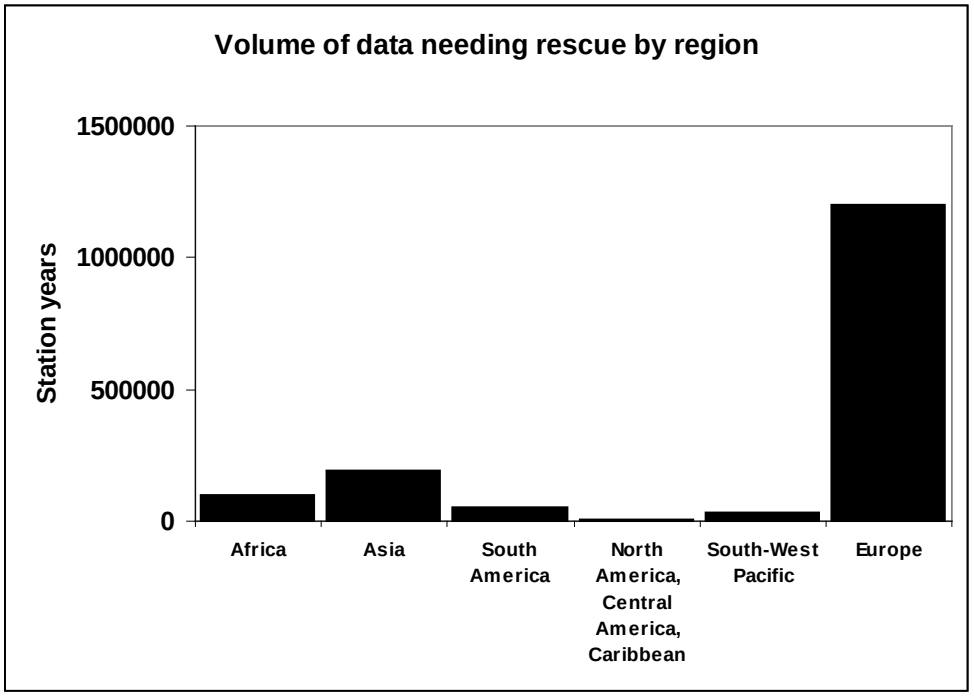
Experience of previous data rescue activities did not seem to determine whether data were secure. The survey revealed that 11 of the 19 countries not declaring a need for data rescue had undergone previous successful data rescue projects, all of which were funded and undertaken internally. A similar proportion (21 of 37) of countries in need of data rescue had experienced previous data rescue activities, 16 of which had been entirely internally funded. However the countries with no current need generally described these past projects as involving comprehensive digitisation of paper records, whereas those declaring a need described a partial digitisation of records, involving principal stations only. Descriptions of the unsuccessful projects portrayed a wide number of reasons for failure: limited funds and personnel; equipment failure; inappropriate database capabilities; problems with conversion; loss of raw data; lack of space for storage of paper records; data stored on redundant media.

Usage of a range of hydrological Database Management Systems (DBMS) was reported. All of the countries not in need of data rescue used a hydrological DBMS for managing data, and all new data was entered into these systems; 84 % of these countries stored gauging station metadata

within the system. Of those countries in need of data rescue, 84 % used a hydrological DBMS, 97% of these entered all new data into the system, and only 61 % of them stored gauging station metadata within the system. A wide range of systems were mentioned, many being proprietary systems developed within the NHS, though the system used was not related to the need for data rescue.

Figure 1 shows the volumes of data described as needing rescue, by region. These numbers can only be seen as broadly indicative, as the information provided was inconsistent and in some cases incomplete. They suggest the volumes of data requiring rescue within Europe are far higher, but this is presumably because of the higher density of river monitoring networks in this region historically.

Figure 1 Volumes of data in need of rescue by region.



The survey requested information concerning the type of data rescue required. Table 2 summarises the results, of the 39 NHS requiring data rescue.

Table 2 Summary of data rescue needs.

Question	Details	Number of countries	Percentage (of the 37 countries in need of data rescue)
Is the volume of data in need of rescue increasing?		26	70 %
What types of data are in need of rescue?	Rainfall	29	78 %
	Water Level	34	92 %
	Discharge	31	84 %
	Snow	11	30 %
	Groundwater	22	59 %
	Water Quality	26	70 %
	Metadata	32	86 %
	Rating	26	70 %
	Other	11	30 %
What are the reasons for data rescue need?	Deteriorating media	32	86 %
	No hardware	16	43 %
In what storage media are the data in need of rescue?	Paper Chart	31	84 %
	Paper Manuscript	29	78 %
	Paper Tape	3	8 %
	Floppy Disk	14	38 %
	Magnetic Tape	9	24 %
	Other	10	27 %
Is the data in need of rescue in one location?		18	49 %
Is there a catalogue / inventory of the data in need of rescue?		18	49 %

These results describe a palpable data rescue need internationally. Whilst the scale of the problem is large, with key datasets such as rainfall and river flows at risk in many countries, the availability of inventories is encouraging. The potential benefits to the international scientific community of these data are large, and there must be significant possibilities for improving understanding of river flow data through the capture of station metadata and rating information.

The variety of media requiring data rescue represents a significant challenge, and could be used as an indicator of the urgency of projects. The window for retrieving data from redundant electronic formats, such as magnetic tape and floppy disk, is limited and could provide a focus for efforts. Paper records do not indicate a large risk in themselves, but deterioration can be rapid. Unfortunately the survey responses regarding the deterioration of media did not differentiate between records held on paper and those on magnetic media. However additional information

regarding the urgency of the data rescue need was requested and indicated a variety of specific issues: paper records at risk due to rodent and termite attack and damage from humidity; potential for loss due to fire; lack of electronic backup facilities. 19 of the 37 countries described the need as urgent, due to risk of data loss. Others suggested there was a need that was either being gradually met, or that the need was for additional information e.g. for policy development, climate change studies and flood modelling.

The data rescue survey was considered successful. The number of responses was high for this type of survey approach, and responses were elicited from a broad range of countries, illuminating the severity of the issue in even the most developed nations. The responses provide good information both for increasing our understanding of the international data rescue need, and for targeting assistance to meet this need. The urgency of data rescue requirements vary, but there is a very clear message concerning an urgent need to rescue documents at risk in 25 % of countries responding to the survey. Follow-up work attempting to access data inventories and to catalogue the specific need in those countries describing a data rescue requirement would be an appropriate step to obtain more detailed information about the scale of effort required to tackle the problem.

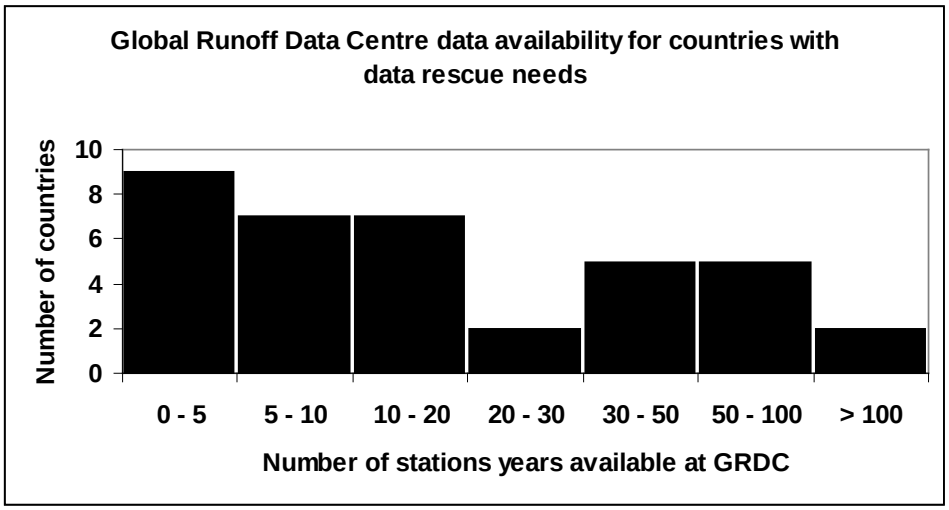
INTERNATIONAL NEED FOR HYDROLOGICAL DATA RESCUE

From the information above, the requirements for data rescue at a national level are well described. However the importance to the international community that hydrological data are preserved is often not understood. Some of the potential benefits could be improved spatial coverage, improved length of record, and improved access to gauging station metadata.

Figure 2 shows the number of station years of daily flow data within the GRDC database for those countries describing a need for data rescue. Whilst many countries stating a need for data rescue are well represented within the GRDC database, for 9 of these countries there are 5 or less station years of daily flow data available, with 16 having 10 years or less. Data rescue within

these countries alone, assuming agreement concerning data sharing, would lead to an important improvement to the international data available. The area of these 9 countries, currently represented within the GRDC by 19 station years of daily flow data, totals 13.1 million km^2.

Figure 2 Data currently available internationally for those countries declaring a data rescue need.



In addition to the benefits of increasing areal coverage of flow measurement, there is potential for increasing the length of records available. However, the hydrological data rescue survey did not ask for information regarding specific stations, or lengths of record and so it is currently not possible to quantify this.

There is an international demand for gauging station metadata. Lack of information pertaining to data quality and anthropogenic impacts fundamentally undermines the use of river flow data for many applications. Svensson et al (2005) describe flood and low flow trend analysis in data from 21 carefully selected GRDC flow series, but highlighted the lack of information stored within the data centre to assess their suitability regarding changes to the stations or within the catchments. A means for capturing this metadata from countries providing flow data is essential. To this end the GRDC have produced a metadata standard (Dornblut, 2009), based on common international standards for monitoring from the Open Geospatial Consortium (OGC). This provides the means

for describing metadata including catchment information, data processing steps, and data quality information. Such standards are welcome but are reliant upon NHS capturing the information and, specifically, having the ability to ensure their data can meet these standards. With appropriate tools for NHS, adequate support, and encouragement for providing, in particular, rating quality information, the utility of a dataset such as the GRDC could be hugely improved for a range of applications, including more accurate understanding of climate change impacts on river flows. FRIEND could play a significant role in improving international hydrological databases by taking the lead in gathering this metadata, and assessing its utility.

CONCLUSIONS

Investigation into the need for hydrological data rescue within National Hydrological Services for the purpose of maintaining national capability and supporting international science has shown that:

- There is a huge need for hydrological data rescue in most countries around the world.
- A quarter of countries have urgent data rescue requirements.
- The majority of gauging station data in need of data rescue is within European NHS, whilst the most urgent need for data rescue is within the regions of Africa, South America, and North America, Central America and the Caribbean.
- Countries needing data rescue are often poorly represented within international river flow databases.
- Most countries wish to rescue gauging station metadata, which is urgently needed to improve the utility of international river flow databases.
- FRIEND could help by including metadata gathering activities within its work on regional hydrological databases.

ACKNOWLEDGEMENTS

The Hydrological Data Rescue survey was coordinated by the World Meteorological Organisation, with the input of Ross James at the Australian Bureau of Meteorology.

REFERENCES

Dornblut, I. (2009). Hydrologic Information – Metadata. Semantic structure for the description of hydrologic data. *Global Runoff Data Centre Report 39.*

Giles, J. (2005) Solving Africa's climate data problem. *Nature.* **435**, 863.

Hirsch, R. M. and Costa, J. E. (2004) US streamflow measurement and data dissemination improve. *EOS*. **85**(20), 197 – 203.

Servat, E. & Demuth, S. (eds) (2006) FRIEND – A Global Perspective 2002 – 2006. Koblenz, Germany.

Svensson, C., Kundzewicz, W. Z. and Maurer, T. (2005) Trend detection in river flow series: 2. Flood and low-flow index series / Détection de tendance dans des séries de débit fluvial: 2. Séries d'indices decrue et d'étiage. *Hydrological Sciences Journal*. **50**(5), 811 – 824.

Vorösmarty, C. J. (2002) Global water assessment and potential contributions from Earth systems science. *Aquat. Sci.* **63**(4). 328 – 351.

WMO (1996) The adequacy of hydrological networks: a global assessment. *Tech. Document No. 740.* World Meteorological Organisation, Geneva.

WMO (2002) Report of the CLICOM-DARE workshop (July 2000) and of the International Data Rescue Meeting (September 2001). T*ech Document No. 1128*. World Meteorological Organisation, Geneva.

World Bank (1993) *Sub-Saharan Africa hydrological assessment: project summary document.* World Bank, Washington DC.