Hydrology and Earth System Sciences, 4(4), 669–679 (2000) © EGS

Appendix
Data management and data archive for the HYREX Programme

A.M. Roberts[1], I.D.Cluckie[2], L. Gray[3], R.J.Griffith[2], A. Lane[2], R.J.Moore[1] and M.A.Pedder[4]

[1] Centre for Ecology and Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK
[2] University of Bristol, Department of Civil Engineering, University of Bristol, Bristol, BS8 1TR, UK
[3] British Atmospheric Data Centre, Rutherford Appleton Laboratory, Chilton, Oxon, OX11 0QX, UK
[4] University of Reading, Department of Meteorology, Reading, RG6 6BB, UK
e-mail for corresponding author: rm@ceh.ac.uk

Abstract

Since the mid 1980s, changes in political imperatives plus technological changes in computer hardware and software have heightened the awareness of the economic value and importance of quality datasets to scientific research. The Natural Environment Research Council's (NERC) interdisciplinary Thematic and Special Topic Programmes have highlighted the need for a coherent data management policy to provide and preserve these quality datasets for posterity. The Hydrological Radar EXperiment (HYREX) Special Topic Programme brought together multi-disciplinary researchers from UK public sector laboratories and universities. In this paper, the HYREX data management strategy, its problems and its solutions are discussed. The HYREX data archive, situated at NERC's British Atmospheric Data Centre, is described.

Keywords: radar, data, archive, web, storm, flood

Introduction

Recent changes in Government policies, scientific research organisations and computer technology have precipitated a fundamental re-evaluation of the importance of good quality databases to scientific research. The UK Government, in its policy paper 'Realising our Potential' (HMSO, 1993), highlighted the need to improve the nation's 'wealth creation and quality of life' by ensuring the efficiency and effectiveness of Government-funded research. In further policy papers, the Government highlighted the continuing importance of sustainable development (HMSO, 1994a) and environmental issues after the Rio Earth Summit (HMSO, 1994b). The Royal Commission on Environmental Pollution concluded that the monitoring of environmental trends is an area of science to which government must provide long-term commitment and financial support.

In addition, the problem of 'cost recovery' for data is now significant. This issue is shaping the sale of data collected and processed by government agencies (Rhind, 1992; Haines-Young, 1992; Haines-Young and Watkins, 1996) and has caused many agencies to re-assess the commercial value of their data. To address the generic problem of data acquisition, data management, data use and charging for data, the UK Natural Environment Research Council (NERC) has implemented a Data Policy Plan (NERC, 1996).

Advance in any area of technology creates pressures for further advances. The increased ability to measure environmental variables has yielded large volumes of data, leading to a need for ever greater processing power. Over the past few years quantum leaps in computer hardware, software and database technology have created further pressures. The implication of having such vast data processing and data storage power has ensured a greater priority for issues of data management in UK scientific research.

In response to these driving forces, NERC has changed the organisation of its research and the Data Policy Plan is raising the profile of data and data systems in UK science. It has become increasingly important for NERC to work across traditional discipline boundaries to align its strategic initiatives with the end user. Hence, NERC has funded Special Topic and Thematic Programmes (NERC, 1995) which bring together communities of scientists from public-funded laboratories and universities to address a particular scientific problem across traditional scientific boundaries. The results from these research Programmes should be applicable to a wider user community and thus improve the UK's 'wealth creation and quality of life'.

This paper outlines how the issues of data awareness, access and quality have been addressed in a NERC Research Programme using the example of the Hydrological Radar EXperiment (HYREX). The paper describes the datasets for the programme, and the associated issues of quality, restrictions of use and methods of access.

The HYREX Programme

Over the past ten years or more, the UK has led the world in research in the use and deployment of ground-based weather radars. These radars are now routinely used to support forecasts of extreme weather events which would lead to river flooding. Despite the research that had already been undertaken, it was obvious to hydrologists and meteorologists that enormous opportunities still existed for fundamental scientific research using weather radar. This belief provided the basis and motivation for initiating HYREX as a NERC Special Topic Programme,

The scientific problem recognised by HYREX was the need for an improved understanding and mathematical description of the movement of water, in both the land and atmospheric phase of the hydrological cycle, in order to make spatial measurements of precipitation. The ability of weather radar to measure precipitation in all its phases, in two and three dimensions, provides an opportunity to progress research in this area.

The HYREX Programme posed five major scientific questions for its multi-disciplinary teams of researchers:

- What are the structures in rainfall systems, and can the parameterisation of mesoscale and cloud physics models be improved by using radar?
- What is the spatial structure of rainfall fields, and can this be established by comparisons of point and spatial estimates of rainfall using dense raingauge networks and radar?
- Which is the most effective radar system for the discrimination of precipitation (rain, snow, hail, graupel) and for the subsequent use of precipitation estimation in hydrological forecasting?
- Can the outputs from global satellite-based rainfall measurements be improved by using radar/rainfall networks for calibration?
- Do improved spatial rainfall estimates from radar lead to better estimates of runoff and is there any scale dependency?

To answer some of these questions, the HYREX management board awarded funding for six projects within the programme. These projects fell into three main areas of research:

- Rainfall measurement. Research into the design of radar/rainfall networks; the verification of polarisation radar techniques for better estimates of rainfall; and radar hydrometeorology using vertically pointing radar.
- Rainfall field modelling. This area of research covered the statistical aspects of spatial-temporal rainfall fields.
- Rainfall forecasting. Stochastic modelling and the development of a space-time rainfall forecasting system for real-time flow forecasting. Research into methods for determining short period precipitation and flow forecasting.

INFRASTRUCTURE OF THE HYREX PROGRAMME

HYREX was a three year Special Topic Research Programme that ran from 1993 to 1996. The focus of the Programme was the Brue catchment in south-west England, extending over 135 km^2 to the river gauging station at Lovington. The catchment rises from 19.8 m at the outfall to a maximum altitude of 244 m. It is 79% grassland, with 10% arable farming, 7% forestry and no major urban or industrialised areas. The catchment is scanned by three radars: a conventional C-band radar at Wardon Hill, a new Doppler C-band radar at Cobbacombe Cross and an experimental Doppler dual-polarisation S-band radar at Chilbolton. A mobile vertical pointing X-band radar was also deployed in the catchment.

A new dense network of 49 raingauges was installed in the Brue catchment as a key part of the HYREX infrastructure. The network mostly has at least one recording gauge in every 2 km square, providing catchment-wide coverage, increasing to two gauges along two NE-SW lines (aligned with the prevailing storm direction and orthogonal to the topography). The network included two sub-networks having 8 gauges in a 2 km square, each arranged in a diamond-within-a-square configuration, with one situated in an area of low relief and one in an area of high relief.

Automatic weather and soil stations, located within the low relief 8 gauge sub-network, were operated continuously, whilst an optical disdrometer, radio sondes and C130 aircraft flights were available for deployment during special Intense Observing Periods (IOPs).

Data Management Strategy

PROGRAMME DATA MANAGEMENT PROBLEM

With the considerations of the NERC Data Policy and the exponential growth in data collection and archiving in mind, a new approach to the management of data was required for NERC Research Programmes. Lowry and co-workers examined the problems of data management and quality control to support Thematic Programmes in the marine sciences (Lowry and Cramer, 1995; Lowry and Loch, 1995). They found that the scale and cost of the science involved had raised the awareness that the datasets produced are a valuable long-term resource and should not be allowed to remain solely in the hands of individual scientists. To justify their high costs, Programmes must produce 'deliverables'. In science these usually take the form of published papers, but a further deliverable is now a clearly identified Programme dataset.

THE DATA MANAGEMENT SOLUTION

NERC, realising the potential for commercial and academic use of its data, has established seven Designated Data Centres (DDC). These are the Antarctic Environmental Data Centre, the British Atmospheric Data Centre (BADC), the British Oceanographic Data Centre, the National Geosciences Information Service, the National Water Archive (NWA), the Environmental Information Centre and the NERC Scientific Services Data Centre. The HYREX management committee identified the BADC as the most appropriate Data Centre for storing and disseminating the HYREX datasets.

BRITISH ATMOSPHERIC DATA CENTRE

The British Atmospheric Data Centre (BADC) – the NERC Designated Data Centre for the atmospheric sciences – is sited at the Rutherford Appleton Laboratory (RAL) in Oxfordshire, part of the Central Laboratory of the Research Councils (CLRC). The role of the BADC is to assist UK atmospheric researchers to locate, access and interpret atmospheric data and to ensure the long-term integrity of atmospheric data produced by NERC projects. In addition to the provision of long-term data archiving facilities, as in the case of the HYREX project, the BADC also provides a service for the acquisition and dissemination of third party datasets that are required by a large section of the UK atmospheric research community and are most efficiently made available through one location. In particular, the BADC has direct access to the Met Office database and is able to negotiate bulk access to meteorological data required by the research community.

BADC datasets are held on-line and are available through the World Wide Web (http://www.badc.rl.ac.uk/). Users can then transfer the specified files over the network to their computer. Software is provided to read and assist in the manipulation of BADC datasets and extensive information is provided on the data collection procedures, formats, data quality, contact names and references to journal papers. Also available through the BADC web pages are descriptions of approximately 30 further datasets that are archived and distributed by the BADC and useful links to over 300 other web sites of interest to atmospheric researchers. A more detailed description of the contents of the HYREX database at BADC and instructions on accessing the data are provided later in the section headed 'Data Availability'.

HYREX Datasets

Of the six projects awarded funding under HYREX, three involved responsibility for collecting and archiving datasets. These three projects are outlined below in terms of the datasets they provided to the HYREX community database.

PROJECT 1: DESIGN OF RADAR/RAINGAUGE NETWORKS FOR HYDROLOGICAL USE

This project, undertaken by the Institute of Hydrology, reviewed the requirements for rainfall field estimates in the hydrological sciences and established how these can best be met using networks of radars and raingauges. It quantified the accuracy of raingauge and radar networks in measuring rainfall at different spatial scales utilising the HYREX dense raingauge network and the continuously scanning network radars over the Brue catchment. A second stage looked at the sensitivity of catchment flow models to rainfall uncertainty and variability utilising the river flow data recorded at Lovington on the River Brue. The project gathered data from the common experimental infrastructure, carried out data quality control, maintained a HYREX community data archive and transmitted the data to the BADC HYREX archive for use by the international scientific community. The Ministry of Agriculture Fisheries and Food supported this data management activity financially.

The main datasets were those collected from the dense raingauge network, the two C-band network radars at Wardon Hill and Cobbacombe Cross and the river gauging station at Lovington on the River Brue. Field collection of raingauge data, using data loggers, was carried out by AllWater Technology under contract to the Environment Agency, who also provided the river flow data (recorded at 15 minute intervals), whilst the radar data were supplied by the Met Office. The raingauges were Cassella 0.2 mm tipping-bucket gauges recording time-of-tip to a time resolution of 10 seconds. The network radars complete one azimuthal scan every minute, and cycle through 4 different scan elevations (0.5, 1.0, 1.5 and 2.5 degrees) every 5 minutes. Software at the radar site converts measurements of reflectivity on a radial grid into measurements of rainfall intensity on two Cartesian grids: an 84×84 grid of 5 km square pixels covering a radius of 210 km and a 76×76 grid of 2 km square pixels covering a radius of 76 km. Rainfall intensity is digitised into 208 levels ranging semi-logarithmically from 1/32 to 126 mm h^{-1}. Only the Cartesian-grid rainfall intensity data are recorded, with data available for all scan elevations on the 5 km grid, but only for the lowest scan ('beam 0') on the 2 km grid. The Brue catchment lies within the 2 km grid for both radars.

The NERC Equipment Pool supplied an automatic weather station recording solar and net radiation, wet and dry bulb temperature, wind speed and direction and rainfall;

it was modified to include measurements of atmospheric pressure, recording at 15 minute intervals and 0.2 mm tip rainfalls. The Institute of Hydrology provided a soil water station providing hourly measurements at three depths from capacitance probe, pressure transducer and temperature sensors, along with 0.5 mm tip rainfall. It also deployed an optical disdrometer during some Intense Observing Periods. This instrument records the number and beam occupation time of raindrops in 16 drop-size 'bins', with a sampling interval of between 10 and 60 seconds. The weather and soil stations and disdrometer were located nearby in Bridge Farm orchard within the 2 km square of the low relief 8 gauge sub-network.

PROJECT 2: HYDROMETEOROLOGICAL STUDIES USING A VERTICALLY POINTING RADAR

This research project, undertaken by the University of Salford, commissioned and deployed a mobile X-band Vertically Pointing Radar (VPR) in the HYREX project area to support a high resolution study into the hydrometerology of local rainfall systems. Of particular significance was an examination of the problems of bright band contamination of quantitative rainfall measurements from scanning weather radars. Large errors can result from the transmitted microwave intersecting water coated ice crystals at, and immediately below, the melting layer and thus generating larger reflected signals than would arise from ordinary raindrops.

The research project was designed to further the understanding of bright band characteristics and dynamics and support the development of algorithms capable of modelling the consequences of variations in the vertical reflectivity profile on quantitative precipitation measurements. Hence, the VPR data were integrated with other HYREX data sources including the Chilbolton, Cobbacombe and Wardon Hill radars as well as the Salford low-cost C-band MARS data (see Cluckie *et al.*, 1995). Additional VPRs (one mobile and one located at the University of Salford) were used to investigate the spatial variability of bright band characteristics through a number of transect experiments in north-west England and via deployment in areas covered by the Chilbolton radar.

Ground truth measurements were obtained from Transportable Weather Stations (TWS) and two drop counting raingauges (Hydragauges) deployed alongside the VPRs. Data were also obtained from an acoustic radar and a disdrometer. Additional rainfall measurements from the dense raingauge network in the Brue catchment were also used.

DATA DESCRIPTION

Vertically Pointing Radars

Three VPRs were used during the course of the project. Technical specifications may be found in Cluckie *et al.* (1998). For the period May to December 1994, data were collected in the north-west of England from one device located at the University of Salford and two mobile devices which were deployed locally either at Warrington and Fazakerley or Davyhulme and Audenshaw (see Tilford *et al.*, 1995). These combinations covered ranges of 43 and 17 km respectively. During 1995 the two mobile VPRs were operated in the HYREX region extending eastward to Chilbolton. Table 1 details the availability of all instrumentation operated by this group during the period of the HYREX programme.

The VPR data consist of vertical reflectivity profiles at a resolution of 7.5 m and 4 secs, and the files can be read using software available via the BADC HYREX web site. The data quality is somewhat variable due to hardware and software modifications made during the course of the project and the unattended operation of the radars. No attempt has been made to quality control these datasets except in the case of specific events analysed by the research group.

Transportable Weather Stations

TWSs were operated alongside each of the VPRs and the availability of the data from these is outlined in Table 1. The stations recorded rainfall (0.1 mm tips), wind speed and direction, pressure, wet and dry bulb temperatures, humidity and net and solar radiation, all with a temporal resolution of 2 minutes.

Hydragauges

The Hydragauge is a high resolution raingauge (Stow, 1993) which produces near-constant-size drops which are counted during a switch-selected time interval to give the rainfall rate. The nominal calibration of 160 drops mm^{-1} enables the gauge to measure rainfall intensities of up to approximately 200 mm h^{-1}.

Disdrometer

An RD-69 disdrometer (Joss and Waldvogel, 1970) was deployed alongside a VPR and the acoustic radar. The device records drop sizes over a range of 0.3–5.0 mm and stores data as the number of drops in each of 20 approximately exponentially distributed drop size classes over the given range. For the purposes of this study, the temporal resolution was set to 1 minute.

Acoustic Radar

The acoustic profiler (Bradley and George, 1994) transmits a beam at 4250 Hz, 4300 Hz or 4350 Hz sine wave for 0.075 s from a 150 W array of 37 piezo-electric tweeters. These are then used as microphones for approximately 2 s which enables the signal reflected from inhomogeneities up to an altitude of 340 m to be observed. The returned signal is amplified and sampled at 10 bit resolution at a rate of 572 Hz over 20 m range gates.

Table 1. VPR dataset data availability

Date	Location	VPR	TWS	Hydragauge	Disdrometer	Acoustic Radar
19.4.95–3.5.95	Chilbolton	✓		✓		
19.4.95–3.5.95	Boscome Down	✓	✓	✓		
4.5.95–24.8.95	Middle Wallop	✓	✓	✓		
4.5.95–10.5.95	Boscome Down	✓	✓	✓	✓	
11.5.95–23.5.95	Boscome Down	✓	✓	✓	✓	✓
24.5.95–05.6.95	Boscome Down	✓	✓	✓	✓	
08.6.95–14.6.95	Boscome Down	✓		✓		
23.6.95–30.6.95	Boscome Down	✓	✓	✓		
1.7.95–26.7.95	Boscome Down		✓	✓	✓	
27.7.95–1.8.95	Boscome Down	✓	✓	✓		✓
2.8.95–9.8.95	Boscome Down		✓	✓		✓
10.8.95–14.8.95	Boscome Down	✓	✓	✓		✓
15.8.95–24.8.95	Boscome Down		✓	✓		✓
25.8.95–20.9.95	Middle Wallop	✓		✓		
25.8.95–20.9.95	Alhampton	✓		✓		
21.9.95–23.11.95	Middle Wallop	✓	✓	✓		
21.9.95–3.10.95	Alhampton		✓	✓		
4.10.95–17.10.95	Alhampton	✓	✓	✓		
18.10.95–21.11.95	Alhampton	✓	✓	✓		
24.11.95–21.12.95	Alhampton	✓		✓		

DATA QUALITY AND RESTRICTIONS OF USE

Data analysis has been restricted to a number of individual storm events and consequently no general quality control measures have been undertaken. The use of any of these datasets should be made with this in mind. Some guidance from the group at the University of Bristol can be given as regards instrument failure and any potential error sources found during the course of routine maintenance. Most of these datasets can be made available to any interested third parties although the use of VPR data by other researchers should be undertaken in collaboration with the originators because of the technical issues detailed earlier.

PROJECT 3: METHODS FOR SHORT-PERIOD PRECIPITATION AND FLOW FORECASTING INCORPORATING RADAR DATA

This project was undertaken as a collaboration between the University of Reading Joint Centre for Mesoscale Meteorology (JCMM) and the Institute of Hydrology (IH). The research effort at Reading focussed on detailed analysis of Mesoscale Atmospheric data obtained during Intensive Observation Periods (IOPs). The main tool used in this research was the Met Office Mesoscale Model (MM), which is a particular configuration of the Unified Model (UM). For each IOP, the model was run with initial data identical to that used in the operational forecast model runs, obtained from the Met Office. The model runs consist of specific case studies of forecast durations between 6 and 24 hours and correspond to the IOP periods. The model was run with and without Observational Data Assimilation to see how the inclusion of extra upper air sounding data affects the moisture budget calculations. The model runs were performed on the Cray YMP-8 at RAL using a portable version of the UM (version 3.4).

One of the aims of the research at Reading was to use the wind and humidity fields from the MM in the moisture budget equation. During the IOPs, estimates of the volume-averaged atmospheric moisture flux convergence affecting southern England were obtained. The area for which the moisture budget is estimated was varied systematically to define the lower limits in space and time for which the moisture budget equation provides a realistic prediction of observed precipitation. In each case, the relative importance of the different terms in the equation was examined.

To verify the moisture budget, areal averages of rainfall were obtained from analysis of raingauge measurements made by IH and then compared with the results from the model. Fields of surface temperature and pressure and estimates of updraught velocity from the MM were used by IH in a water-balance storm model for short-term rainfall and flow forecasting, with radar used to update the atmospheric water content.

DATA DESCRIPTION

Mesoscale Model Variables

The UK operational Mesoscale model uses a rotated latitude and longitude coordinate system in which the computational north pole is shifted to an actual position of 37.5 N, 177.5 E. The corners of the computational area are approximately in actual latitude/longitude: (60.1N, 16.6W; 60.2N, 10.7E; 46.6N, 12.7W; 46.7N, 7.1E). The grid-length is 0.15° in each direction which is approximately 16.8 km, giving 92 by 92 grid points running from the north-west corner. In the vertical the MM uses hybrid sigma/pressure coordinates and has 31 levels with the first 28 of these being moist levels.

Output data from a model run are written to 3 output files ('PP' files) which contain single level and multi-level fields. The multi-level fields are either held on model levels or on pressure levels. The largest dimension of a field data volume is 92 by 92 by 31. Documentation regarding contents of the output PP files is shown in Table 2.

The Time Profiles are described below:

Table 2. Mesoscale Model data fields

Field Name	Time Profile	Domain Profile
Pstar after time step (ts)	T15MI	DS
U compnt of wind after time step (ts)	T15MI	DM
V compnt of wind after time step (ts)	T15MI	DM
Theta after time step (ts)	T15MI	DM
Specific Humidity after time step (ts)	T15MI	DMW
Surface Temperature after time step (ts)	T15MI	DS
10 metre wind U-comp	T15M	DS
10 metre wind V-comp	T15MI	DS
Evap from soil surface -amount kg m^{-2} ts^{-1}	T15MA	DS
Evap from canopy - amount kg m^{-2} ts^{-1}	T15M	DS
Evaporation from sea (GBM) kg m^{-2} s^{-1}	T15MA	DS
Surface Latent Heat Flux w m^{-2}	T15MI	DS
Temperature at 1.5m	T15MI	DS
Specific Humidity at 1.5m kg m^{-2} ts^{-1}	T15MI	DS
Temperature after large scale precip	T15MI	DM
Large scale rain amount kg m^{-2} ts^{-1} ts^{-1}	T15MA	DS
Large scale snow amount	T15MA	DS
Large scale rainfall rate kg m^{-2} s^{-1}	T15MI	DS
Large scale snowfall rate kg m^{-2} s^{-1}	T15MI	DS
Cloud liquid water after ls precip	T15MI	DMW
Cloud ice content after ls precip	T15MI	DMW
Convective rain amount kg m^{-2} ts^{-1}	T15MA	DS
Convective snow amount kg m^{-2} ts^{-1}	T15MA	DS
Convective rainfall rate kg m^{-2} s^{-1}	T15MI	DS
Convective snowfall rate kg m^{-2} s^{-1}	T15MI	DS
Pressure at convective cloud base	T15MI	DS
Pressure at convective cloud top	T15MI	DS
Temperature after convection	T15MI	DM
Omega on model levels	T15MI	DM
U component of wind on pressure levels	T15MI	DP
V component of wind on pressure levels	T15MI	DP
T on pressure levels U grid.	T15MI	DP
Omega on pressure levels U grid	T15MI	DP
Specific Humidity, P level U grid	T15MI	DP
Geopotential height: pressure levels	T15MI	DP
Temperature on pressure levels	T15MI	DP
Pressure at mean sea levels	T15MI	DS
Geopotential ht of model levels	T15MI	DM

- T15MI–Indicates a field output at time-steps separated by 15 minutes.
- T15MA–Indicates an accumulated field output at time-steps separated by 15 minutes, with an accumulation interval of 5 minutes (i.e. every time-step).

The Domain Profiles are described below:

- DS–Indicates a Single level field.
- DM–Indicates a field output on atmospheric model levels, over the range of model levels from: 1 (bottom level), to: 31 (top level).
- DMW–Indicates a field output on atmospheric model levels, over a range of model levels from: 1 (bottom level), to: 28 (top wet level).
- DP- Indicates a field output on the following Pressure levels (in hPa): 1000, 950, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250.

MESOSCALE MODEL DATA ARCHIVE

Mesoscale Model data from selected IOPs and other periods of interest are archived at BADC. Data from 27 model runs for 5 IOPs and some other selected events have been archived.

DATA QUALITY AND RESTRICTIONS OF USE

Quality of the model run data is not applicable here and there are no restrictions on the use of these data.

Data availability

REVIEW OF USER REQUIREMENTS

The requirements of a long-term data management system are that:

- the data are adequately archived,
- they are adequately documented, and
- that they are readily accessible.

The first requirement involves maintaining the integrity of the chosen storage media. It requires ensuring that multiple copies of the data are kept in separate locations in case of hazards, such as fire or flooding, and may require periodic transcription of the data onto alternative media. It also requires that the data are stored in formats that are likely to be readable for the foreseeable future. For example, where possible, ASCII formats rather than binaries are preferred. Ensuring the integrity of valuable data in these various ways is one of the central missions of the NERC Data Policy and hence of each of the NERC Dedicated Data Centres.

The second requirement is particularly important and one that is often overlooked. It is essential that adequate documentation accompanies the data so that the data are still usable even when the expert who collected them has moved on to other projects and is no longer available. This information is often referred to as 'metadata' i.e. information about the data. Data on their own without adequate documentation of how they were acquired, the prevailing conditions under which they were taken, the format in which they were recorded and a statement on the quality of these data, are virtually useless. It is also important that the level of detail provided in the documentation is adequate. The documentation should be aimed not only at the immediate HYREX research community but also at the wider research community who may not be fully conversant with the measurement instruments or their interpretation. Adequate documentation has been an essential prerequisite for the acceptance of any dataset as part of the authorised long-term HYREX data archive. Various levels of documentation are available through the BADC WWW pages (http://www.badc.rl.ac.uk/). They range from a brief overview of the HYREX project, its aims and achievements through to the very detailed documentation accompanying each instrument dataset.

The final requirement, not least in importance, is that the data are readily accessible. The first hurdle for a potential new user of a dataset, even before the user is able to decide whether it is an appropriate dataset to use, is simply to find out what data are available and what will be the investment required to make effective use of the data. The provision of good documentation that is easily accessible plus a user-friendly system for data access, usually in the form of a data catalogue, is the key to ensuring that potential users do not give up at this early stage. The following sections describe the facilities that have been put in place for access to the long-term HYREX data archive.

DATA AVAILABILITY THROUGH THE BADC

The policy for HYREX data management is to provide a single point of entry to the long-term archive for new users of the data and this is provided by the BADC WWW pages (http://www.badc.rl.ac.uk/data/hyrex/). These pages provide an overview of the project and of the data that are available, which include raingauge data, radar data, automatic weather station (AWS) data, automatic soil water station (ASWS) data, river level and flow data, disdrometer data, land use data and output from special runs of the Met Office Unified Model (UM). The total data holdings available through BADC amount to approximately 7.6 Gigabytes, of which 5.4 Gigabytes are data from the Unified Model. In cases where there are no restrictions on access to the data, the data files themselves may be accessed directly via the BADC WWW pages. In the case of restricted datasets, detailed instructions on how to gain access to the data are provided. This will usually involve logging directly on to the BADC computer. Methods for bulk extraction of data for the more experienced user are also available. The

BADC also acts as the long-term archive of documentation for the datasets and in all cases detailed 'help files' describing the data access methods, data formats, available software etc. are provided through these web pages.

THE BADC HYREX CATALOGUE

The diverse nature of the various datasets makes them rather confusing for a potential new user. In order to provide both new and experienced users with the ability to identify quickly the availability of required data, a catalogue of the HYREX datasets has been developed. The user interface to the catalogue is a BADC WWW page with various search options that can be constrained as the user requires. The options include the data type, instrument type, instrument location, time period of measurement and data sampling resolution. When the required search restrictions have been chosen, a catalogue query is then initiated by the user. The response to this query will be a list of data filenames that satisfy the search criteria together with a description of various other attributes of the files. A null response indicates that no data satisfying those criteria are available. The user may further restrict and repeat the query until satisfied that the optimum files have been located. The returned list of filenames will also act as links to the data files themselves, provided there are no restrictions to public access of the data through the WWW pages. The user can then proceed to click on a filename in the list and immediately view and transfer the contents of that file. In the case of restricted datasets, the returned list of file names provides adequate information that the user may locate easily and extract the file(s) once he has logged on to the BADC computer.

DATA ACCESS THROUGH THE BADC

Raingauge, Soil Water, Disdrometer, Land-Use and Automatic Weather Station Data

These data are not restricted and are therefore available through the BADC WWW pages by navigating to the relevant pages and thence the required data directories. The standard WWW utilities may then be used to transfer the file to the user's local computer. The data are organised in a series of sub-directories: /badc/hyrex/data/nn where nn is one of the instrument types provided in Table 3 e.g. the automatic soil water station data are found under the directory /badc/hyrex/data/asws/.

The individual data files are located under the directories indicated in Table 3, together with appropriate on-line documentation describing the file-naming conventions, the specific contents of each of the file types and the file formats for that particular instrument.

The data are stored as ASCII files, the majority of which have been compressed using the standard UNIX compress utility to reduce the total volumes and for ease of transfer across the network. Once uncompressed, the files are plain text, in most cases a series of single-line entries, with data items in each entry separated by commas or spaces. Table 4 provides more details of the time span and time resolution of each file, typical file sizes and the total dataset size from each instrument.

Table 3. HYREX directory structure

asws/	Automatic soil water station data
aws/	Automatic weather station data
dis/	Disdrometer data
drn/	Dense raingauge network data
landuse/	Landsat-derived land use data
rgs/	River gauging station data
cob/	Cobbacombe radar data
war/	Wardon Hill radar data

Cobbacombe and Wardon Hill Radar Data

Data from Cobbacombe and Wardon Hill radars are available for *bona fide* research purposes only. Prospective users of these data are therefore required to sign an agreement to this effect before access to the data is provided. Copies of this agreement are available from the BADC WWW pages and may be transferred and printed locally. Users need to sign and return the agreement to the BADC and they will be provided with an account on the BADC computer to access the restricted sub-directories. The organisation of the restricted sub-directories is identical to the unrestricted data directories described in the previous section. File transfer protocol (ftp) is the most appropriate method for transferring these files to the user's local computer.

Chilbolton Radar Data

Data from the Chilbolton radar are held on CD-ROM, copies of which can be requested either from the BADC or directly from the Radio Communications Research Unit (RCRU) at the Rutherford Appleton Laboratory who also have WWW information on this data product (http://rcru1.te.rl.ac.uk/). A directory of sample gif files of radar images for certain measurement periods is provided as an illustration of the data. These can be viewed directly through the web or pulled over the network for plotting. Typical sizes of the gif files are 60-85 kbytes. For further data manipulation and for data collected during other periods the data are available on a set of 6 CD-ROMs. These data are in binary form and descriptions of the file formats are available through the RCRU WWW pages. Software is also available for reading, displaying and manipulating the data.

Unified Model Data

30 Gbytes of model data are available through the BADC

Table 4. BADC HYREX Datasets

Time coverage of dataset	Time coverage of each file	Time resolution of data	File compressed	Typical file size (Kb)	Approx. total data size (Kb)
Automatic Soil Water Station data					
20.9.94-18.11.96	1 (part) year	hourly	yes	70-270	563
Automatic Weather Station data (two types)					
'daily' data					
01.9.93-18.11.96	1 (part) year	daily	yes	3-11	32
'minute' data					
01.9.93-18.11.96	1 (part) year	15 min	yes	100-800	2,000
Disdrometer data					
14.2.95-15.2.95					
10.7.95-12.7.95					
25.7.95-26.7.95					
6.9.95-11.9.95					
6.10.95-6.10.95	1 hour	10 or 60 secs	yes	0-15	577
Dense Raingauge Network data					
1.9.93-24.10.96	1 (part) month	10 secs	yes	24-250	5,400
Landuse data					
N/A	N/A	N/A	no	650	650
River Gauging Station data					
1.4.95-31.9.96	1 (part) year	15 min	yes	200-250	2,600
Chilbolton Radar data					
8.4.94 - 12.3.96	1 day	15 min	no	1-10	3,000,000
Cobbacombe Radar Data					
1.2.94-30.9.96	1 day	5-15 min	yes	<2000	808,000
Wardon Hill Radar data					
1.9.93-30.9.96	1 day	5-15 min	yes	<2000	1,300,000
Brue Catchment Area Averaged data (spatial and time averages)					
Spatial average gauge data	1 month	15 min	yes	12	439
Cobbacombe radar (2km res.)	1 month	15 min	no	90	2,600
Cobbacombe radar (5km res.)	1 month	15 min	no	90	2,100
Wardon Hill radar (2km res.)	1 month	15 min	no	90	2,600
Wardon Hill radar (5km res.)	1 month	15 min	no	90	2,100
Daily averages of above 5 datasets	3 (part) years	daily	no	18	92
UK Meteorological Office Unified Model Data					
19.10.94	daily	15 min	yes	80,000 to 480,000	30,000,000
05.12.94					
06.12.94					
14.02.95					
10.07.95					
16.07.95					
07.06.96					

from model integrations corresponding to the Intensive Observation Periods. Output from 24 model integrations is available for the dates shown in Table 4. For each date there are various model integrations available, for example, with and without assimilation of extra HYREX radiosonde data. For each model run, three files are produced containing a total of 39 data types. Some of these are single-level fields e.g. surface temperature, 10 m winds, *etc.*, while some are multi-level fields available on both 31 model levels and 16 pressure levels. The files are stored in 32 bit IEEE binary format and file sizes vary between 80 Mbytes and 480 Mbytes. For a full list and description of the model runs available the reader is referred to the BADC WWW pages.

Due to the large data quantities, the model dataset is physically located on a robotic tapestore and so the data are effectively near-line rather than on-line. Because of this, it is not possible to provide access to the data files through the BADC WWW interface or through anonymous ftp access. The data are accessed by logging on to the BADC computer and running the retrieval software provided to request access to the appropriate robotic tape identified using the HYREX catalogue. The complete contents of the tape are then transferred to a disc on the BADC computer and the user may then extract the data files required. The retrieval software is portable and runs under both UNIX and VMS. More experienced users may wish to install it on their local computers and, hence, access the robotic tapestore directly rather than through the BADC computer. Software to read and extract subsets of the data is also available and consists of a FORTRAN program run by a UNIX shell script. The software has been tested on a SUN and a CRAY. The software will also run under DEC UNIX provided the correct compiling options are used. Details of this are provided in help files available under the WWW. Various routines are also available for manipulation and plotting of the model data.

Conclusion

The HYREX Programme has demonstrated the benefit to community science of creating a valuable dataset that can be shared by the participants during the research programme and be made available to the international research community for subsequent research. Web technology has been used to create a metadatabase containing information on the HYREX data archive and a mechanism for passive dissemination of the datasets, via remote FTP access, to a worldwide research community. This investment in infrastructure will ensure that the science of weather radar and its use in hydrology can advance through unfettered access to the HYREX datasets by *bona fide* researchers.

Since the formal closure of the NERC Special Topic Programme in 1996 it has been possible to extend the database further. The National Rivers Authority and its successor the Environment Agency, recognizing the benefits of longer term datasets, funded the continuation of the dense raingauge network over the Brue through to Spring 2000. The Institute of Hydrology continued with quality control and archiving of the data and maintained the weather and soil stations over this period. Radar data from Wardon Hill and Cobbacombe for this extended period are being added to the archive. When completed in 2001, this extended archive will be transferred to the BADC web site for use by the international research community.

Acknowledgements

HYREX was funded by the Natural Environment Research Council (NERC) as a three year Special Topic Research Programme running from 1993 to 1996. The impetus for this Programme arose from recommendations from NERC and the UK's Interagency Committee on the Hydrological Use of Weather Radar and had the full support of the National Rivers Authority (NRA), the Met Office, Ministry of Agriculture, Fisheries and Food (MAFF), North West Water Ltd. and the Rutherford Appleton Laboratory (RAL). NERC Council awarded grants to the research groups involved in the HYREX Programme, while the above agencies provided infrastructure to support the HYREX community. The NRA provided a dense raingauge network and river flow data from the Programme's experimental catchment; AllWater Technology, under contract to the NRA, installed and maintained the raingauge network and provided the data to the Institute of Hydrology (IH) for quality control and archiving. The Met Office provided radar data from Wardon Hill and Cobbacombe, radio sondes and C130 flights. MAFF funded a HYREX community database at IH. North West Water Ltd. provided funding for a Vertically-Pointing Radar (VPR) operated by the University of Salford. Doppler dual-polarisation radar data from the Chilbolton radar were provided by RAL, the NERC Equipment Pool provided an Automatic Weather Station and IH supplied a Soil Water Station plus optical disdrometer.

References

Bradley, S.G. and George, K.R., 1994. The use of a small acoustic wind profiler from a sea-going platform. *Int. J. Remote Sensing*, **15**, 237–244.

Cluckie, I.D., Tilford, K.A. and Tyson, J.M., 1998. Vertical reflectivity profiling using a high resolution X–band radar. In *Advances in hydrological applications of weather radar*, G.W. Shepherd and H.-R. Verworn (Eds.), Proc. 2nd Int. Symp. on Hydrological Applications of Weather Radar, Hannover Germany, September 1992, Ch. 16, 132–140.

Haines-Young, R.H., 1992. The Tradable Information Initiative, *Geographical Information 1992. Year Book of the Association for Geographic Information*, I. Heywood and J.C. Hudson (Eds.), Taylor and Francis, London, 231–238.

Haines-Young, R. and Watkins, C., 1996. The rural data infrastructure. *Int. J. Geogr. Information Systems*, **10**, 21–46.

HMSO, 1993. *Realising our potential: a strategy for science, engineering and technology*. HMSO, 74pp.

HMSO, 1994a. *Sustainable development: the UK strategy*. Cm 2426, HMSO, London.

HMSO, 1994b. *Biodiversity: The UK Action Plan*. Cm 2428, HMSO, London.

Joss, J. and Waldvogel, A., 1970. *Distrometer RD-69. Instruction Manual*, Distromet Ltd, PO Box 33, CH-4020, Basel, Switzerland, 32pp.

Lowry, R.K. and Cramer, R.N., 1995. Database applications supporting Community Research Projects in NERC marine

sciences. *Geological Data Management*, Geological Society Special Publication 97, 103–107.
Lowry, R.K. and Loch, S.G., 1995. Transfer and SERPLO: powerful data quality control tools developed by the British Oceanographic Data Centre. *Geological Data Management*, Geological Society Special Publication 97, 109–115.
NERC, 1996. *NERC data policy handbook*. Version 1.0, Natural Environment Research Council, Swindon, UK, 20pp.
NERC, 1995. *1995 Corporate Plan Update*, Natural Environment Research Council, Swindon, UK, 24pp.
Rhind, D., 1992. Data access, charging and copyright and their implications for geographical information systems. *Int. J. Geogr. Information Systems*, 6, 13–30.
Stow, C.D., 1993. *Hydra sensor manual*. Internal Report, Dept. of Physics, University of Auckland, New Zealand, 12pp.
Tilford, K.A., Han, D. and Cluckie, I.D., 1995. Vertically pointing and urban weather radars. In *Hydrological uses of weather radar*, K.A. Tilford (Ed.), British Hydrological Society Occasional Paper No. 5, 147–164.

Hydrology and Earth System Sciences, 4(4), *681* (2000) © EGS

Hydrology & Earth System Sciences

Author Index to Volume 4