LEGACY ATMOSPHERIC SOUNDING DATASET PROJECT

BY RICHARD H. JOHNSON, STEVEN F. WILLIAMS, AND PAUL E. CIESIELSKI

For over a half-century, field campaigns have played a central role in advancing the atmospheric sciences. Although recent decades have witnessed organized efforts at cataloging and archiving field campaign data for both United States and international experiments, largely through the auspices of the National Center for Atmospheric Research (NCAR), data from campaigns from the 1950s into the 1980s have not been systematically collected and archived at a central location. Here we report on an effort to take steps to correct this situation, with an initial focus on atmospheric sounding data. Furthermore, we make an appeal to the international community to assist in this effort by identifying and locating observations from past field campaigns and reporting this information to our project website.

FIG. 1. A photo depicting some of the archival media used to store data for many of the early field experiments. Plans are to extract data from such media and place them in a common format at a central archive.

Invariably, field campaigns involve substantial resources to conduct special or intensive observations over selected locations around the globe. The resulting data have been invaluable as the main observational bases for advancing tropical, midlatitude, and polar research, much of it involving processes associated with clouds and precipitation systems. Examples of such past field campaigns are the 1969 Barbados Oceanographic and Meteorological Experiment (BOMEX); the 1974 GARP Atlantic Tropical Experiment (GATE); the 1985 Oklahoma-Kansas Preliminary Experiment for STORM-Central (OK PRE-STORM); and the 1992–93 Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE).

These field campaign datasets have been collected and used by various organizations and research groups around the world. Their archival status varies from one field program to another since no consistent data management strategy has been applied over this long period of time. Figure 1 shows some of the archival media (e.g., 9-track magnetic tapes, microfiche, paper printouts, etc.) on which these earlier datasets are stored. Those who wish to access these data generally have to contact a variety of places to locate and obtain the data files, and in many cases

AFFILIATIONS: JOHNSON AND CIESIELSKI—Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado; WILLIAMS—National Center for Atmospheric Research, Boulder, Colorado
CORRESPONDING AUTHOR: Richard H. Johnson, Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523
E-mail: Johnson@atmos.colostate.edu

DOI:10.1175/BAMS-D-11-00092.1

it is not clear if the data even still exist, especially for older campaigns. For most field programs in the past two decades, the NCAR Earth Observing Laboratory (EOL) maintains a comprehensive online data-archive and access system. However, with the passage of time there is a danger that some of the older datasets will be lost or languish on obsolete media that are difficult to read, and that "corporate memory" of such datasets by principal investigators and project participants will slowly fade away. Moreover, these datasets represent one of the main legacies of field programs, and the observations are taken at great investment of time and resources. Preserving their legacy should therefore be a high priority of both the funding agencies and the scientific community.

One component of field program datasets that has a particularly long-term value to the scientific community is the atmospheric vertical profile represented by upper-air sounding data. The sounding networks for several past field campaigns in the tropics and midlatitudes are shown in Fig. 2. Sounding observations from these networks have numerous applications: description of environmental conditions, which provide a context for validating and understanding other observations (radar, satellite, etc.); diagnostic studies of heat and moisture budgets from which the properties of convection can be inferred; computation of advective properties of heat and moisture to force cloud-resolving models and single-column models, which aid in the improvement of parameterization schemes; initialization of models for reanalyses and case studies; and model validation. Upper-air datasets from field programs are generally of higher quality than operational soundings because the large suite of instruments deployed in these field campaigns allows for cross-calibration that can greatly enhance data accuracy. Given the special efforts invested in producing high-quality sounding measurements in various locations around the world over the past five decades, there is also great value from the standpoint of climate applications in preserving the data.

Unfortunately, there is currently no central location for access of all research-quality sounding data from past national and international field experiments. As previously mentioned, NCAR, as well as the National Climatic Data Center (NCDC), has holdings from several more recent experiments, but collections from older experiments are incomplete or not available. Due to the usefulness and importance of these sounding datasets, and considering the enormous investment in terms of time, effort, and financial resources to collect them, a collaborative data stewardship effort involving scientists at NCAR EOL and Colorado State University is underway with the following objectives:

1) Identify past field programs for which central collections of sounding data do not exist.
2) Track down existing holdings of sounding data for those field programs, to the extent they exist, at centers, laboratories, and universities.
3) Extract sounding data that are found from old storage media (i.e., 9-track tapes, printouts, etc.), and place them into a consistent, common digital format.
4) Carry out standard quality control of the sounding data, including objective gross limit and vertical consistency checks.
5) Prepare a catalog and a central, publicly accessible archive of the sounding data (at NCAR and/or NCDC).

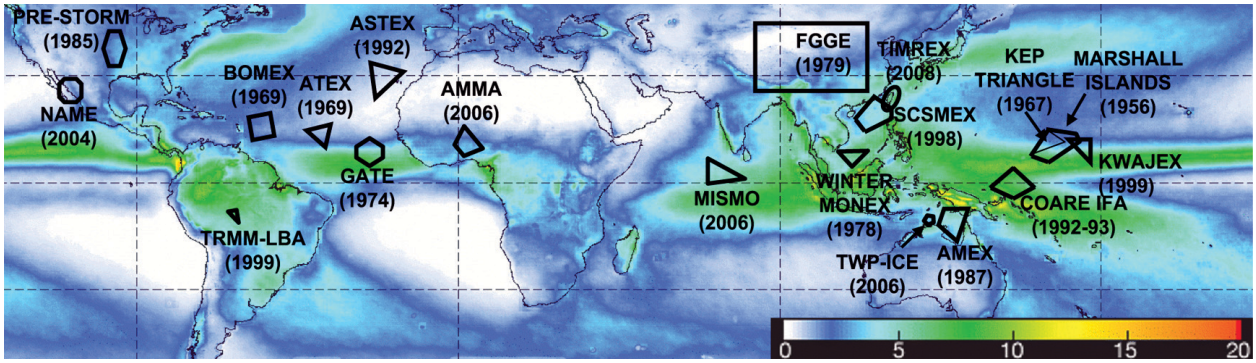

Fig. 2. Sounding networks for several major field campaigns superimposed on the annual-mean TRMM 3B43 rainfall (mm/day) for the period 1998–2008.

Experiment Name	Beginning Date	Ending Date	PIs	Host Countries	Location	Upper-Air Sondes	Surface	Aircraft	Radar	Ocean	Other
Marshall Is.	04/1958	07/1958	Yanai	US	West Pacific	?	-	-	-	-	-
AVE 1	02/1964	02/1964	-	US	SE US	digital	-	-	-	-	-
Line Is.	02/1967	04/1967	Zipser	US	Central Pacific	digital	-	digital	-	-	-
FSU Barbados	07/1968	09/1968	Garstang	US	Barbados	?	?	?	-	?	?
ATEX	02/1969	02/1969	Augstein	Germany	Central Atlantic	?	-	-	-	-	-
BOMEX	05/1969	07/1969	Holland, Kuettner	US	West Atlantic	digital	-	-	-	-	-
ISMEX	05/1973	07/1973	-	India, USSR	India	digital	-	-	-	-	-
GATE	06/1974	09/1974	Kuettner	US, USSR	East Atlantic	digital	-	-	-	-	-
AMTEX	01/1975	02/1975	Lenschow	US	NW Pacific	?	?	?	-	-	-
MONSOON77	05/1977	08/1977	-	India, USSR	India	digital	-	-	-	-	-
Winter MONEX	12/1978	02/1979	Krishnamurti, Webster, Houze	-	Malaysia	-	-	-	-	-	-

Fig. 3. Top portion of field-program inventory table posted on Legacy Atmospheric Sounding Data Set web page showing relevant information for each experiment. Under the data columns, a "?" symbol indicates that the dataset is known to exist but has not yet been located. The entry "digital" means the dataset has been located in a digital form, whereas "digital coded" means the dataset currently exist in a digital form that we have not yet been able to read (e.g., binary EBCDIC). A blue entry under "Experiment Name" indicates that a separate web page exists for this field program.

Because of uncertainty in the location and availability of the older archives and data tapes, the above tasks have been divided into two phases: tasks 1 and 2 will be accomplished under Phase 1, and tasks 3 through 5 in a proposed Phase 2.

PHASE 1: IDENTIFICATION AND LOCATION OF SOUNDING DATASETS. Current efforts in Phase 1 of this project have identified 51 field programs with sounding datasets. Of these 51, sounding datasets in some form (digital, paper, tapes, etc.) have been located for 39 of them. Of these 39 datasets, 34 have data in an easily readable form. To help us locate missing datasets and potentially other field program datasets we may have overlooked, we recently established a Legacy Sounding Data Set website (www.eol.ucar.edu/projects/legacy). To summarize our activity, this website contains a table (Fig. 3) listing the various field programs and their pertinent information (program name, dates, project PIs, host counties, and an indicator of which datasets for the field program have been located). While we are initially focusing on sounding datasets, other field datasets (e.g., radar, aircraft, surface, fluxes, profiler, etc.) associated with the sounding-intensive field campaigns will be identified and considered for future archiving as they are located. While outside the scope of this initial project, we plan to eventually identify, prioritize, and archive these additional datasets from not only the 51 field programs but also for others for which such archives are not yet established. As part of Phase 1, we are requesting the assistance of the atmospheric science community in helping us locate these missing sounding datasets.

PHASE 2: SOUNDING PROCESSING AND ARCHIVAL. As sounding datasets are located, we will identify those which require further recovery (e.g., reading magnetic tapes, writing software to read coded formats, scanning in printed output, etc.). Our proposed goal is to place each dataset into an easily readable form. Once this is accomplished, the sounding datasets will be processed with a software package developed by NCAR/EOL called Atmospheric Sounding Processing Environment (ASPEN), which does some limited quality control and removes bad data values. If resources are available, a user-friendly version of each dataset will be created by interpolating the high vertical resolution data to a uniform vertical resolution (e.g., 5-hPa) and assigning quality flags to each data value.

EOL will track results from the data recovery effort using its internal Data Tracking System (DTS) and create online project data access web pages through the EOL Data Management System (EMDAC) for each field program. The summary table on the legacy website will be continually updated and links to web pages for the major field programs will be provided. Several examples of web pages for past field campaigns can be viewed by referring to the table at www.eol.ucar.edu/projects/legacy. These web pages will contain a description of the experiment, along with access to the data and supporting documentation. Appropriate metadata will be created for these datasets that will allow data discovery through other interoperable distributed data archives.

SUMMARY. Observations from field campaigns over the past 50 years have led to significant advances in our understanding of many weather phenomena. However, with the passage of time a number of the older field datasets may become unrecoverable for a variety of reasons. Over the course of the past two years, we have embarked on a project to identify, locate, and archive field campaign sounding datasets, which have considerable value to the atmospheric science community. While reasonable progress has been made, a number of key datasets have yet to be located. In this regard, we are soliciting the assistance of the atmospheric science community to help us locate these missing datasets. *If you are aware of sounding data sources from older field campaigns for which our holdings are incomplete (see the table at* www.eol.ucar.edu/projects/legacy*), we ask that you contact us via this web link*. Additional details regarding this project can be found at the website. Once the project is closer to completion, an article will be prepared to document the effort and notify the community about the availability of the data.

ACKNOWLEDGMENTS. This project has been supported by the NOAA Climate Projects Office with grant NA090OAR4320074 to Colorado State University and National Science Foundation (NSF) grant ATM0856145 to NCAR EOL. Special thanks go to Jay Fein, Eric DeWeaver, and Anjuli Bamzai at NSF and Chris Miller and William Murray at NOAA for their support and encouragement in pursuing this project. We thank Scot Loehrer and Linda Cully for their assistance in setting up the Legacy Atmospheric Sounding Data Set Web page and many of the field experiment web pages, and Brian McNoldy for his help in creating the sounding network figure. Comments from Ed Zipser and two anonymous reviewers are much appreciated.