Conf-920163--1

PNL-SA--19880

DE92 006732

CLOUDS AND RADIATION TESTBED DATA ENVIRONMENT: SITE DATA SYSTEM AND EXPERIMENT CENTER

R. B. Melton
A. P. Campbell
D. M. Edwards
J. L. Tichler
S. F. Bobrowski
J. M. Corbet
P. Kanciruk

January, 1992

Presented at the
8th International Conference on
Interactive Information and Processing
Systems for Meteorology, and Oceanography,
and Hydrology
January 5-10, 1992
Atlanta, Georgia

Work supported by
the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

CLOUDS AND RADIATION TESTBED DATA ENVIRONMENT: SITE DATA SYSTEM AND EXPERIMENT CENTER

Ronald B. Melton, Stephen F. Bobrowski
Pacific Northwest Laboratory, Richland, WA

A. Peter Campbell
Argonne National Laboratory, Argonne, IL

Jonathan M. Corbet
National Center for Atmospheric Research, Boulder, CO

Donna M. Edwards
Sandia National Laboratories, Livermore, CA

Paul Kanciruk
Oak Ridge National Laboratory, Oak Ridge, TN

Joyce L. Tichler
Brookhaven National Laboratory, Upton, NY

1. INTRODUCTION

The Department of Energy (DOE) has initiated the Atmospheric Radiation Measurement (ARM) program as a research effort to reduce the uncertainties found in general circulation and other models due to the effects of clouds and solar radiation (DOE 1990, Patrinos, et al. 1990). This program will provide an experimental testbed for the study of important atmospheric effects, particularly cloud and radiative processes, and testing of parameterizations of the processes for use in atmospheric models. The design of the testbed, known as the Clouds and Radiation Testbed (CART), calls for five long-term field data collection sites as well as a mobile set of instrumentation to be used in short-term field campaigns. The first of the sites is expected to begin operation in April of 1992.

Within the ARM Program, an experiment has been defined as the prospective test of a model, i.e., the test of a model's predictive capability. An experiment is specified by identifying the model or models to be tested, the model input requirements, the measurements needed for comparison to model outputs, and the measurements needed to diagnose model performance. The identification of required measurements includes the specification of data fusion or other techniques to be used in converting the basic instrument observations into the required set of measurements.

The CART Data Environment (CDE) is the element of the testbed which acquires the basic observations from the instruments and processes them to meet the measurement requirements of ARM experiments. This system will handle approximately 1.2 Gbytes/day of instrument observations when the first site becomes operational. The initial instrumentation for the site will include 50 and 915 Mhz wind profilers and RASS (radio acoustic sounding system), an array of whole sky imagers, a ceilometer, CLASS type sonde system, a network of PAM-type (portable automated mesonet) stations, wide band solar and IR sensors, a surface flux station, a high resolution spectrometer, special wide-band solar sensors, and an instrumented 60 meter tower. The data from these instruments must be quality checked, archived, and processed continuously for use in the experiments of the ARM Science Team. A preliminary summary of the requirements for this system is presented in Melton, et al. (1991).

A formal design process has been used to develop a description of the logical requirements for the CDE. We have grouped these requirements into subsets that are used to drive a parallel software design and implementation process. A key element of this process is the identification of existing software solutions that meet some or all of the requirements and that can be incrementally evolved into a complete solution. For example, we have identified the Zeb software developed at the National Center for Atmospheric Research as a well developed starting point that meets initial data ingest, storage, analysis, display and control requirements for our site data system and experiment center. In parallel with software development, our requirements have been mapped into a high level physical architecture that provides the basis for hardware design.

This paper discusses the overall architecture of the CDE and describes our implementation strategy. It will then focus on the use of Zeb as a starting point for implementation of the site data system and experiment center as examples of the implementation strategy.

2. OVERALL ARCHITECTURE OF THE CART DATA ENVIRONMENT

The CDE has three major physical elements: the site data system, the experiment control center, and the ARM data archive center (see Figure 1). The functional requirements for each of these elements is summarized below.

2.1 Site Data System

The site data system is concerned with:

- Control and operation of instruments
- Acquisition of data from instruments
- Data conversion and calibration
- Quality assessment and flagging of individual data streams
- Data intercomparison in support of instrument operation
- Retaining a record of information about the instruments and their operation
- Logging of site operations activity
- Retaining the record of information about the site
- Data display and analysis for site operators, the site scientist, and the site program manager

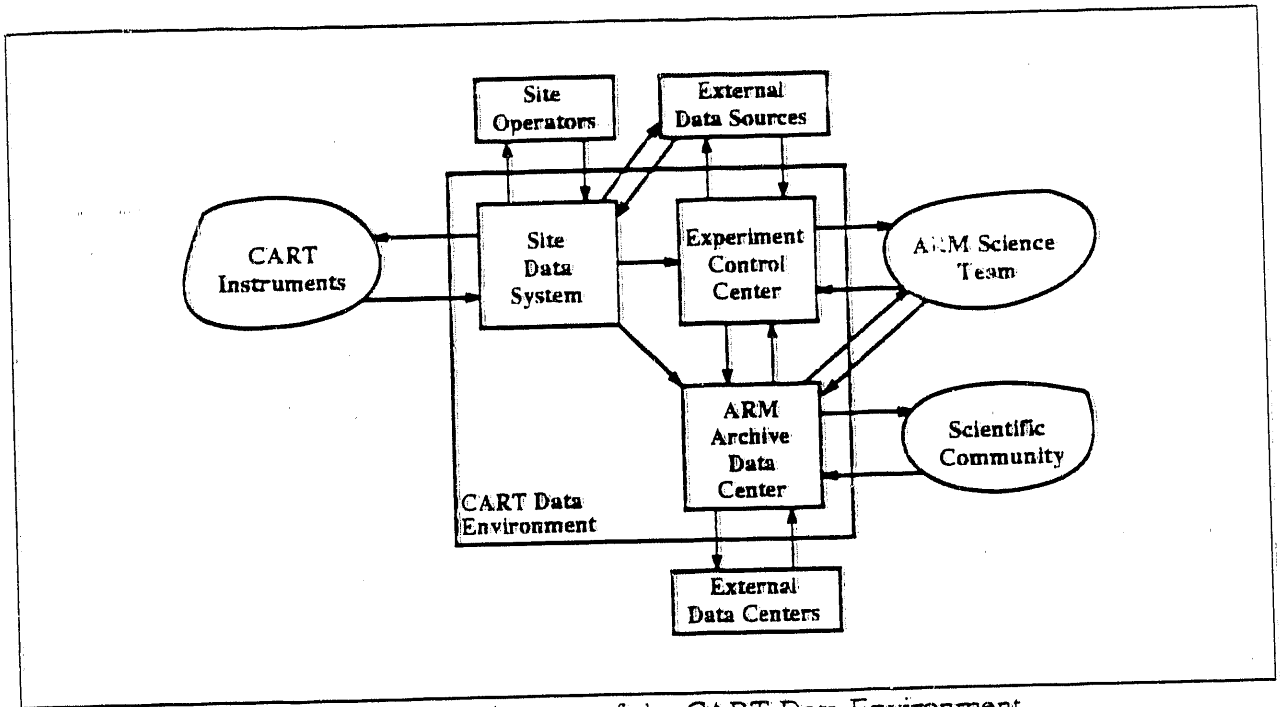

Figure 1: High level physical architecture of the CART Data Environment

The site program manager is an individual from a Department of Energy Laboratory charged with overall management responsibility for the site. The site scientist is an individual responsible for making day-to-day scientific decisions affecting site operations. The site operator is the organization responsible for the operation of the site.

2.2 Experiment Control Center

The experiment control center is concerned with:

- Design of ARM experiments for the ARM Science Team
  - Model inputs
  - Data for comparison with model outputs
  - Data for diagnosing model performance
- Acquisition of non-CART data, e.g. GOES, required to support ARM experiments
- Data intercomparison for data quality assessment
- Generation of data for the active set of experiments
- Delivery of experiment data sets to the ARM Science Team

2.3 ARM Data Archive

The ARM data archive is concerned with:

- Long term storage of CART data
- Value added processing of CART data

- Providing CART data retrospectively to the ARM Science Team and the general scientific community
- Linkage with other data centers

3. IMPLEMENTATION STRATEGY

Our implementation strategy is based on use of existing solutions wherever possible, evolutionary implementation of functionality, and parallel implementation of independent subsystems. This strategy is based in part on principles set forth by Gilb (1988). The variation on his approach comes in pursuing parallel implementation paths.

Our parallel implementation is organized around the physical architecture of the CDE. The major implementation activities are focussed on the site data system, the experiment control center, and the data archive. These implementation activities are focussed on software development through evolution of existing solutions. A fourth implementation activity is focused on integration of the software and, design of the hardware architecture, and on data communications.

We have identified an existing solution, Zeb, that we are using as the starting point for both the site data system and the experiment control center. The next section of the paper describes this software and how we are planning to use it.

4. Zeb: USE OF AN EXISTING SOLUTION

The Zeb system has been developed by the National Center for Atmospheric Research (NCAR) to support atmospheric sciences field campaigns (Corbet and Mueller, 1991). Zeb has a modular architecture that allows new functionality to be incorporated to meet the varying needs of individual campaigns (see Figure 2). The elements of Zeb are: a set of data ingest modules, a data store, a set of modules for processing data, and a set of display and control modules. Each of these elements is briefly described below.

The Zeb data ingest capability is based on ingest modules for each instrument or class of instruments, depending on data volume. For example, radar ingest involves high volumes of data and generally requires an individual module for each radar. Ingest from PAM stations, on the other hand, is a low data volume operation. It is possible, then, to ingest from the entire PAM network with a single ingest module. Zeb has been designed for distributed operation over a network, so the ingest processes can be executed on dedicated resources if necessary. The ingest process formats the data for storage in the Zeb data store.

The Zeb data store uses standard data formats, such as NetCDF (Rew and Davis, 1990). Applications refer to data in the data store by platform or other source of the data. The data store attempts to hide, to the greatest extent possible, the differences between different sources of data, so that display and analysis modules are not required to deal with platform-specific details. It also handles the movement of data between stations on the network, and manages the use of disk space, cleaning up old data when necessary.

4.2 Zeb and the Experiment Control Center

Zeb also meets a substantial portion of the requirements for the experiment control center. The data ingest capability meets the requirement for acquiring non-CART data. The ability to incorporate processing modules meets the requirements for intercomparison of data and for processes that operate on the data to produce data sets required for ARM experiments (for example data fusion or data assimilation processes.)

The requirements not addressed by Zeb are the design of experiments and the logging of experiments. Again, the flexible nature of Zeb will allow for the integration of solutions to these problems into the overall system.

4.3 Zeb and the Data Archive

The relationship between Zeb and the ARM data archive is not well defined at this point. We are currently assessing the general requirements for the ARM Data Archive. Over the course of the ARM Program we expect the archive to grow to 100's of terabytes. Our current plans are to provide access to the archive through abstractions of the data and by identifying views of the data that are of particular interest to archive users. There is great potential for Zeb to be used as a tool for processing data from the archive in creating these custom views. For archive users that have Zeb, data may be distributed as Zeb data store copies.

5. SUMMARY AND CONCLUSIONS

The CDE is a distributed data system that must continuously process data from a variety of instruments. The system must provide for operator control of the instruments and for a variety of processing (e.g. data fusion) that provides data to be used in experiments of the ARM Science Team. Over time the number of instruments and experiments is expected to grow.

The implementation strategy for the CDE is to work from existing software that meets some or all of the requirements for subsystems of the CDE. We have identified the Zeb software as meeting a significant number of the requirements for the CDE's site data system, experiment control center, and possibly some of the requirements for the data archive. We have made the decision to use Zeb as the core of the CDE.

By using Zeb we are realizing several advantages. Some data ingest modules already exist and can be used with minimal modification. Others require extensive modification to be capable of collecting metadata about the instruments. Early testing is enabled for these modules which provides the design team with early experience in implementing the site data system, and for the development of new data ingest modules.

6. ACKNOWLEDGMENTS

This work was supported by the Environmental Sciences Division of the U.S. Department of Energy.

REFERENCES

Corbet, J.M. and C. Mueller. 1991. "Zeb: Software for Data Integration, Display, and Analysis". Proceedings of the 25th International AMS Conference on Radar Meteorology, Paris, France, June 1991. American Meteorological Society, Boston, Massachusetts.

Department of Energy (DOE). 1990. "Atmospheric Radiation Measurement Program Plan". DOE/ER-0441, U.S. Department of Energy, Washington, D.C.

Gilb, T. 1988. Principles of Software Engineering Management. Addison-Wesley, Reading, Mass.

Melton, R.B., et. al. 1991. " Design of the CART Data System for the U.S. Department of Energy's ARM Program". Proceeding of the Seventh International Conference on Interactive Information and Processing System for Meteorology, Oceanography, and Hydrology. January 14 - 18, 1991. New Orleans, LA. American Meteorological Society, Boston, Mass.

Rew, R.K. and G.P. Davis. 1990. "NetCDF: An Interface for Scientific Data Access". IEEE Computer Graphics and Applications, July, 1990, -- 76-82.

Patrinos, A.A., et. al. 1990. "The Department of Energy Initiative on Atmospheric Radiation Measurements: A Study of Radiation Forcing and Feedbacks". Proceedings of the Seventh Conference on Atmospheric Radiation. July 23-27, 1990. San Francisco, CA. American Meteorological Society, Boston, Mass.

END

DATE FILMED

3/23/92