D

Conf-910108--3

PNL-SA--18385

DE91 006722

JAN 28 1991

DESIGN OF THE CART DATA SYSTEM FOR THE U.S. DEPARTMENT OF ENERGY'S ARM PROGRAM

January 1991

R. B. Melton
A. P. Campbell
D. M. Edwards
P. Kanciruk
J. L. Tichler

Presented at the
Seventh International Conference on
Interactive Information and Processing
Systems for Meteorology, Oceanography
and Hydrology
January 13-18, 1991
New Orleans, Louisiana

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

Pacific Northwest Laboratory
Richland, Washington 99352

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

PNL-SA-18385

DESIGN OF THE CART DATA SYSTEM FOR THE U.S. DEPARTMENT OF ENERGY'S ARM PROGRAM

Ronald B. Melton
Pacific Northwest Laboratory, Richland, Washington

A. Peter Campbell
Argonne National Laboratory, Argonne, Illinois

Donna M. Edwards
Sandia National Laboratories, Livermore, California

Paul Kanciruk
Oak Ridge National Laboratory, Oak Ridge, Tennessee

Joyce L. Tichler
Brookhaven National Laboratory, Upton, New York

1. INTRODUCTION

The Department of Energy (DOE) has initiated a major atmospheric research effort to reduce the uncertainties found in general circulation and other models due to the effects of clouds and radiation. The Atmospheric Radiation Measurement Program (ARM) is described in detail in the ARM Program Plan (DOE 1990) and summarized by Patrinos, et al. (1990). The objective of the ARM Program is to provide an experimental testbed for the study of important atmospheric effects, particularly cloud and radiative processes, and testing parameterizations of the processes for use in atmospheric models. This experimental testbed, known as the Clouds and Radiation Testbed (CART), will include a complex data system, the CART Data Environment (CDE). The major functions of the CDE will be to

- acquire measurements from instruments and external data sources
- perform quality assessments of the data streams
- create data streams of known quality to be used as model input and compared to model output
- execute the models and capture their predictions
- make data streams associated with model tests available to ARM investigators in near real-time.

The CDE will also be expected to capture ancillary information ("meta-data") associated with the data streams, provide data management facilities for design of ARM experiments (i.e., tests of a given parameterization's predictive ability), and provide for archival data storage.

A formal top-down process is being used to create the functional design of the CART Data Environment. A key component of this process is the use of facilitated design sessions with representative CDE users. This process is structured to maintain a clean separation of functionality and implementation. This separation is important in designing a flexible, extensible, and robust system.

In parallel with the functional design activities, we are conducting technology assessments by developing prototype solutions to specific functional problems and learning from the experience of other research programs. This technical information will be used to map the functional design into a detailed design for implementation.

The first section of this paper presents background information on CART. Next the process for the functional design of the system is described, the functional requirements summarized, and the conceptual architecture of the CDE is presented. Finally, the status of the CDE design activities is summarized, and major technical challenges are discussed.

2. OVERVIEW OF THE CLOUDS AND RADIATION TESTBED

CART will be the major operational element of the ARM Program. CART will consist of four to six permanent base sites chosen based on their climatological significance and ability to support systematic exploration of the performance of cloud parameterization and cloud formation models under a wide range of significant conditions. Figure 1 shows an artist's conception of an ARM site. Each site has three major components: a central facility, a three-dimensional mapping network, and an extended observing network.

Two classes of instrumentation will be located in the central facility: those for measuring the radiation field directly and those intended to characterize the local radiative circumstances, such as surface and cloud properties. A portion of the CART Data Environment (CDE) known as the site data system also will be located at the central facility.

The three-dimensional mapping network will be made up of auxiliary instrumentation stations within a 20-km radius of the central facility. Instrumentation in these stations will be designed to measure the three-dimensional structure of the atmosphere near the central facility and will use fundamental profiling equipment, as well as basic radiometric and meteorological equipment. Data from this instrumentation will be used to reconstruct the cloud geometry surrounding the central facility.

The extended observing network surrounds the mapping network and the central facility. Sixteen to twenty-five extended observing stations will be distributed within the approximately 200 x 200 km area of the extended site. This area

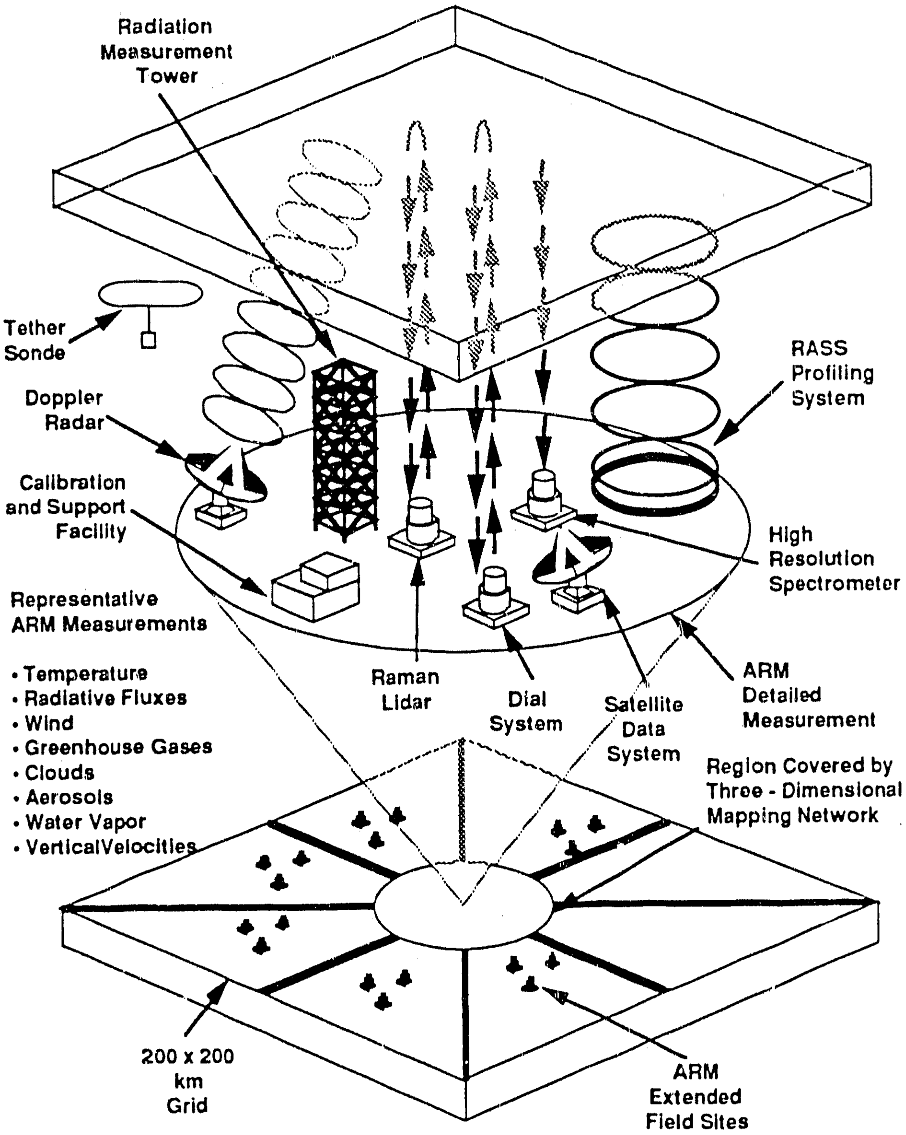

Fig. 1. The ARM instrumentation configuration. An ARM site will have three components as described in the text. The central facility, for which representative equipment is shown, will be supported by a system for mapping the three-dimensional distribution of meteorological data. In addition, 16 to 25 sets of instrumentation will provide critical data for understanding how to generalize the results to the 200 x 200 km GCM grid size.

encompasses a region of the order of magnitude expected for General Circulation Model (GCM) grid cells in the near future. The instrumentation at these stations will be designed to collect basic radiometric and meteorological data. This data will support the development and study of methods used to generalize detailed atmospheric models for use in GCMs and related models through the process of parameterization.

In addition to the instrumentation at permanent ARM sites, ARM will maintain a mobile version of the basic instrumentation found at the central facility along with additional instrumentation for use in directed campaign studies. Aircraft and satellite observations will also be used in ARM.

The data system at each site will collect and process the measurement data streams generated by the various instrumentation packages described above. The data system must produce data streams of known quality that can be used to test the hypothesis that a given parameterization or model is accurate in its predictions. The testing of such hypotheses is the definition of an experiment within the context of the ARM Program.

The data systems at the ARM sites will be linked with each other and with a central data archive. The remainder of this paper addresses the determination of the functional requirements for the CDE focusing on the first ARM site and the data archive. The intention is to stagger implementation of the sites. The design of the CDE will evolve in time as lessons are learned in implementing the site data systems.

3. SUMMARY OF THE FUNCTIONAL DESIGN ACTIVITIES

The functional design of the CART Data System is driven by the functional requirements of the ARM investigators and the other components of CART: Sites, Models, and Instruments. As shown in Figure 2, the requirements for these components are derived from the experimental activities that will be carried out at the CART sites. The data system requirements are purposely shown at the end of the chain of requirements. The data system is intended to be an enabling element that is responsive to the scientific activities of the ARM Program.

We have chosen a systematic approach to the design and implementation of the CART Data System. This approach starts with a thorough definition of required system functionality, translation of functionality into a conceptual or logical system architecture, mapping of the logical system architecture into a physical architecture/design, and implementation of the physical architecture. This process is intended to be iterative and evolutionary. Earlier steps are revisited as our understanding of the functionality of the system matures and detail is developed. An up-to-date description of the functional and logical definition of the system will be maintained throughout its life. This will facilitate system modifications and the design of data systems for CART sites subsequent to the first site.

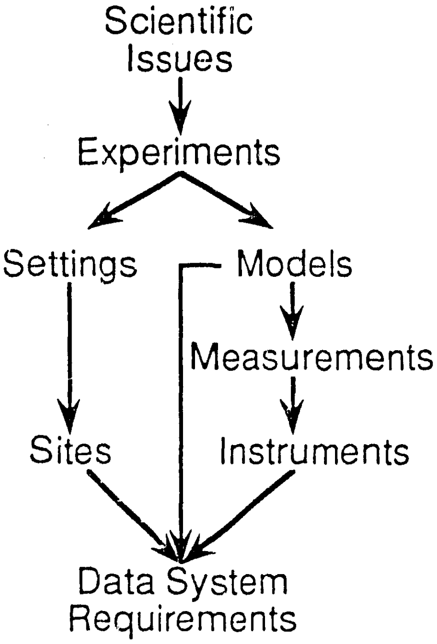

Fig. 2. Functional requirements for the CART Data Environment are driven by other compnents of the ARM Program.

A methodology developed by the WISDM Corporation is being used to complete the functional definition of the CART Data Environment. This methodology has three major components: the use of trained facilitators to work with the design team to identify requirements, a generic system architecture, and a systematic process for identifying requirements using the generic architecture. The remainder of this section describes these components in more detail, particularly the process for identifying the requirements using the generic architecture.

The WISDM® methodology is based on the definition of a system by the users of the system and those responsible for the objects towards which the system directs action. Rather than having a requirements definition and analysis team that conducts interviews with the users and then compiles the results into a functional requirements definition, the WISDM® methodology uses facilitated team sessions. The individuals chosen to participate in the team session represent the various categories of users of the system or those responsible for system objects and are empowered to act in their interest. The team also includes one or more facilitators. The facilitators are neutral third parties whose task is to extract information from the team using a combination of brainstorming and analysis. The facilitators must also understand the general system architecture developed by WISDM® and the agenda for building the definition of a system using the architecture.

The general system architecture used by WISDM® is shown in Figure 3. The architecture is built around the view that a system is defined based on its interactions with its external environment. As shown, the external environment consists of those entities with which the system directly interacts, i.e., those entities it supplies output to, receives input from, or both. Entities that do not directly interact with the system, **A** and **G**, are outside the scope of consideration in defining the functional behavior required of the system.

The agenda for developing a system design works its way from the outside in. The first step is to identify those objects the system is responsible for, i.e., those towards which it directs some action. For information systems this is often the set of objects the system will provide views of to the external environment.

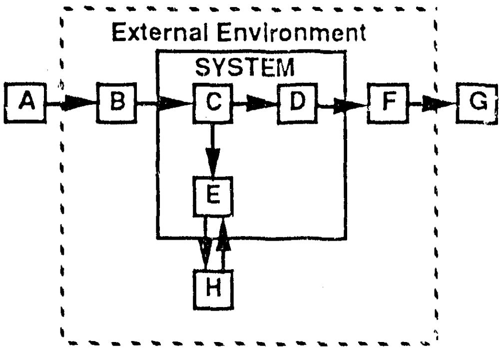

Fig. 3. The WISDM® general system architecture. The system is defined by identifying and analyzing interactions with its external environment as described in the text.

The second step is to define the external entities with which the system will interact (**B**, **F**, and **H** in the figure). This begins the definition of the external environment of the system. It should be noted that a preliminary identification of the system objects and external entities is required to identify the individuals that should participate in the design sessions.

The next step in the process is to describe the external functions that are being performed by the external entities when they are interacting with the system. This is then followed by identifying what information (or other form of output) they require to perform those functions.

Having identified all system outputs the information required to produce those outputs and its sources must be determined. This identifies all external inputs and internal data. For internal data items the ultimate source of the information must also be identified. In other words, all internal data must be directly or indirectly related to one or more external inputs.

At this point the complete external interface model of the system has been defined. Attention now turns to the internal functions of the system, i.e., the producers of internal data and external outputs. The internal functions are implied by the external interface model. One of the rules of modeling for the WISDM® methodology is that all internal functions can produce only one output. Using the external interface model an internal function is identified for each external output or internal data item. The process is completed by naming the internal functions (**C**, **D**, and **E** in Figure 3).

This gives a complete model of the system. At each step above the items (i.e., the external entities, external functions, external inputs and outputs, internal data, and internal functions) are all documented by the team. In addition, general system requirements and future system requirements are identified and documented.

4. CONCEPTUAL ARCHITECTURE OF THE CDE

The conceptual architecture of the CART Data Environment is illustrated in Figure 4. Shown on the left are the sources of measurements that can feed into the data system at a CART site. The primary source of measurements is the CART instruments located at that site. These may be supplemented by observations from satellites (such as GOES) or data from external data sources such as the National Weather Service.

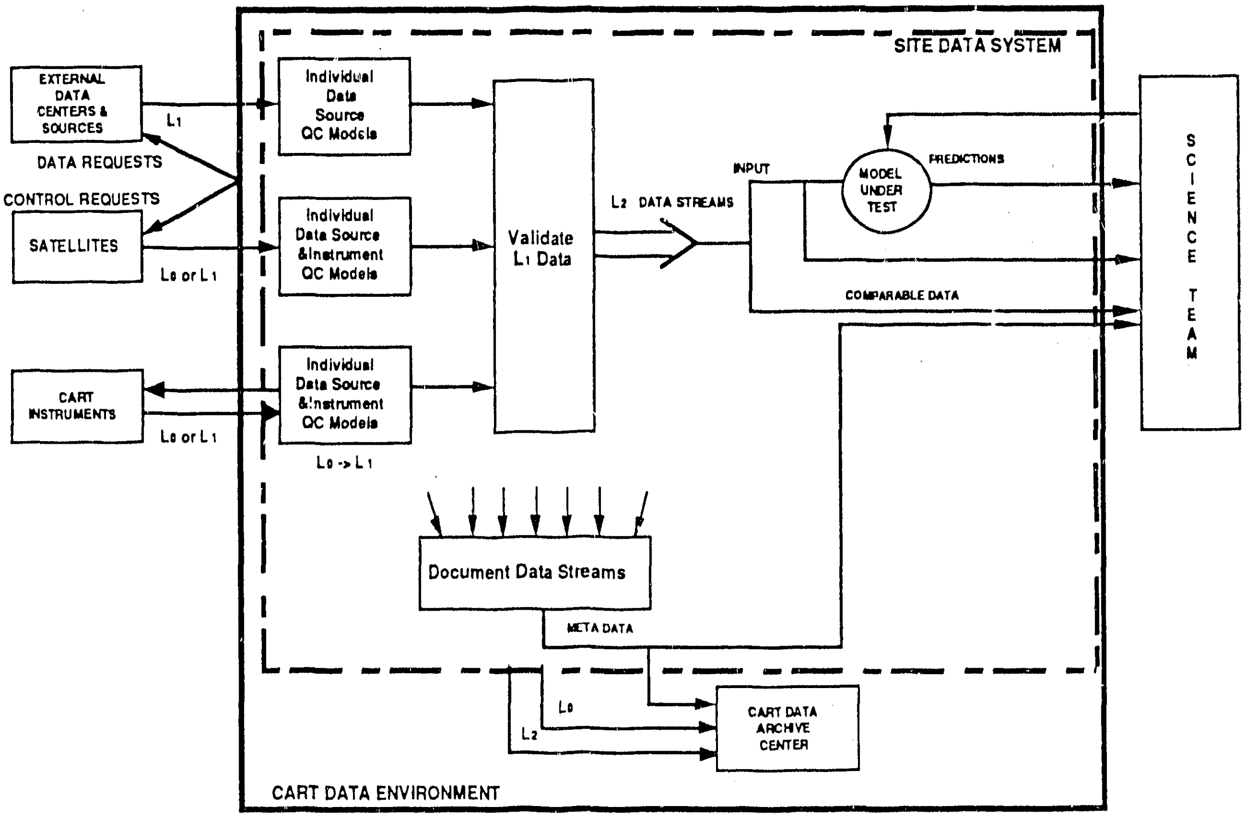

Fig. 4. The conceptual architecture of the CART Data Environment. Data moves from left to right in the diagram. Two quality control steps are used to produce data streams of known quality for use in testing parameterizations or other models. Though testing of only one model is shown, in actual operation many models will be tested simultaneously. In addition, meta-data is automatically generated for the data streams.

Arrows from the data system to the CART instruments and other data sources represent the instrument control signals, data requests or other output required to get data from the external source.

Data coming into the site data system is expected to be either raw instrument signals (L_0 data) or data that has been converted to engineering units and subjected to quality control relative to the performance of the instrument that produced the data (L_1 data). Upon entering the site data system, these data streams will be subjected to a first level of quality control and when necessary converted to engineering units. This will bring the L_0 data to quality level L_1 and verify the quality of data already at the L_1 level. The objective with L_1 data is to estimate the uncertainty associated with each data stream at any point in time.

L_1 data is then subjected to a second level of quality control. In this step all measurements of a given physical parameter are considered together, e.g., all measurements from which temperature can be derived. Two things happen with the data. First, the uncertainties of the data are compared and a decision is made about how to produce the best estimate of the physical parameter. Second, the estimate of the physical parameter is produced either directly from the engineering units of a chosen L_1 data stream or indirectly by combining information in multiple L_1 data streams. The result of this level of quality assessment is L_2 data.

The L_2 data streams are then used as inputs to models and to test the models' predictions (the comparable data). The site data system will deliver the data streams associated with a given experiment to the member of the ARM Science Team conducting the experiment. The data streams delivered include the input to the model, the model predictions, the comparable data, and the meta-data for the three data streams.

The meta-data consists of the information about the data streams, i.e. the documentation of the data streams. The meta-data will be produced simultaneously with the data streams being processed by the data system. It will provide information about which instruments produced the measurements used to estimate the physical parameters, how L_1 data streams were combined, what the uncertainties are estimated to be, and so forth.

In addition to delivering the associated data to members of the ARM Science Team, the site data system must deliver the L_0, L_2, and meta-data to a CART Data Archive Center. The data archive center will be responsible for further documentation of the data selected for long-term storage and for making the data available to the general scientific community.

5. SUMMARY AND FUTURE PLANS

The CDE will be distributed among the CART sites and the CART Central Data Archive. The current estimate is that each site will acquire on the order of 7 Gbytes of data per day under normal operations. The operational goal is to have the data processed to the L_2 level within 15 minutes of its acquisition. At this point only the conceptual architecture of the system that will accomplish these goals has been defined. The next stage of design activities will be to develop a detailed logical design of the CDE by extending the functional design. The primary mechanism for doing this will be to describe each CDE object and its attributes in detail and to define each view of the objects that is delivered as an external output or used as internal data.

The CDE will be implemented in an evolutionary manner. The minimally functional system will be identified from the logical model. This will be mapped to a physical design and implemented first. Lessons learned in this initial implementation will be used to modify the complete system. Further implementation will build around the core system incrementally adding capability.

The nature of the system and these operational goals present several major technical challenges. Three of particular importance are on-line, real-time data quality control, model encapsulation, and automatic meta-data generation. These problems are currently being explored and tested through the implementation of prototype systems both in the laboratory and in collaboration with other field measurement programs. We are also beginning discussion with other atmospheric science programs. We hope to learn from their mistakes and successes.

6. ACKNOWLEDGEMENTS

This work was supported by the Atmospheric and Climate Research Division of the U.S. Department of Energy. We wish to acknowledge Blair Burner and Marcia Hansen of the WISDM Corporation. The material in section three of this paper summarizes the system design process they have developed over the past 10 years. WISDM® is a registered trademark of the WISDM Corporation.

REFERENCES

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

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

03 / 01 / 91