BNL-108010-2015-IR

DEPARTMENT OF ENERGY'S ARM CLIMATE RESEARCH FACILITY EXTERNAL DATA CENTER OPERATIONS PLAN LOCATED AT BROOKHAVEN NATIONAL LABORATORY

Alice Cialella, Laurie Gregory, Kathy Lazar, Min Liang, Lynn Ma, Alison Tilp, and Richard Wagener

Informal Report

Prepared for Jimmy Voyles
ARM Chief Operating Officer
September 19, 2014

May 2015

Biological, Environmental & Climate Sciences Dept.
Brookhaven National Laboratory

U.S. Department of Energy
Office of Science

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Contents

1 Scope
2 Introduction
3 Description
4 XDC Processing Environment
    4.1 Production
    4.2 Development
5 XDC Data Flow Operations
    5.1 Identifying Data for Science Needs
    5.2 Collection
    5.3 Processing
        5.3.1 External data and related VAPs
        5.3.2 Reprocessing and Back-processing
    5.4 Quality Assurance
    5.5 Storing and Archiving
    5.6 Documentation
    5.7 General XDC Data Operations
6 ARM Infrastructure Support
    6.1 Database Support
        6.1.1 Data Quality Databases
            6.1.1.1 Data Quality Reports (DQR)
            6.1.1.2 Data Quality Assessments (DQA)
            6.1.1.3 Data Quality Problem Reports (DQPR)
        6.1.2 Infrastructure Databases
            6.1.2.1 People database
            6.1.2.2 Site Access Requests (SAR)
            6.1.2.3 Metadata Management
        6.1.3 Scientific Databases
            6.1.3.1 Poster database
            6.1.3.2 Research Highlights
            6.1.3.3 Publications
    6.2 Web Server Support
        6.2.1 Web Site Services
        6.2.2 Web Applications
        6.2.3 Web Authentication
        6.2.4 Web Content Management Database Administration
        6.2.5 Web Search Engine
    6.3 Product Registration and Data Delivery
        6.3.1 Field Campaign Data Products
        6.3.2 PI Products
        6.3.3 Value Added Products
    6.4 ARM Ingest and VAP Development
    6.5 ARM Site Surface Characterization
    6.6 ARM Metadata Management
7 Leadership Roles
8 Instrument Mentorship
9 Summary
10 Appendix A: XDC Roles and Responsibilities Table

ARM External Data Center (XDC) Operations Plan

1 Scope

The External Data Center (XDC) Operations Plan describes the activities performed to manage the XDC, located at Brookhaven National Laboratory (BNL), for the Department of Energy's Atmospheric Radiation Measurement (ARM) Climate Research Facility. It includes all ARM infrastructure activities performed by the Data Management and Software Engineering Group (DMSE) at BNL. This plan establishes a baseline of expectation within the ARM Operations Management for the group managing the XDC.

2 Introduction

The XDC was formed in 1996 to manage and document a class of ARM data known as "external data". It is managed by the Data Management and Software Engineering Group (DMSE) (formerly the Scientific Information Systems Group) at BNL.

In addition to the XDC operations, the DMSE has been involved in other ARM infrastructure activities since the early days of ARM. It maintains a number of ARM production databases, some as old as the ARM program. The group has hosted and maintained the ARM Web Server since 1999. In 2001, the XDC assumed responsibility for Field Campaign (FC or IOP) and later Principal Investigator (PI) provided data products. The number of databases has expanded over the years as well.

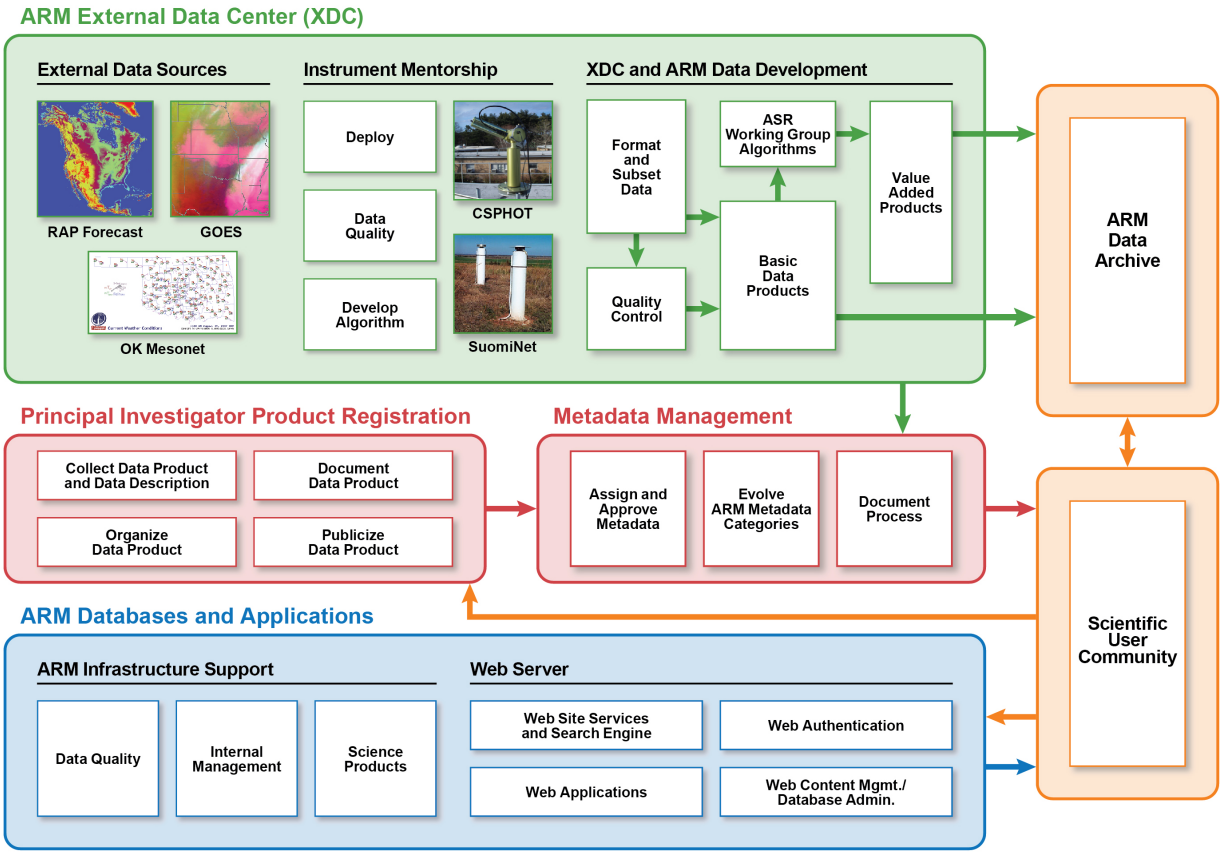
This document begins with a description of the XDC components and is followed by data flow operations, PI product data management, and general operations tasks. The appendices describe the DMSE staff, their roles and responsibilities, their relationships to other parts of ARM management, and problem resolution flows.

3 Description

The XDC collects, documents, and delivers external data of known and quanitfiable quality to the ARM Archive in a timely fashion. External data originate from sources outside of the ARM Program and augment data collected at the ARM sites. Decisions about which data to collect are made with guidance from the working groups which will ultimately result in an Engineering Change Order. The XDC data operations run automatically 7 days a week, 24 hours a day, except where manual intervention is needed, e.g. when the automated processes fail. The XDC staff work normal office hours in the US Eastern Time Zone.

The XDC hosts a number of ARM information infrastructure production databases and ARM and Atmospheric Systems Research (ASR) Web servers.

The XDC also uses a public access server dedicated to the delivery of data by ASR funded (and non-funded) Principal Investigators and Field Campaign participants. XDC staff document, assign metadata and organize the provided data for retrieval on the online Field Campaign archive.



4 Figure 1. DMSE Activities and XDC Operations Flow

4 XDC Processing Environment

The XDC data processing, database and web-server environments consist of two main systems each, a development and production system, with identical computer hardware, operating systems, and software release (major) versions. All software is developed and tested on development before installation on production. In an emergency, e.g. catastrophic failure of the production system, the development system can be reconfigured to run the production processes and services.

4.1 Production

The XDC processes all ARM external data, from collection through delivery to the Archive. The complete life cycle process includes identifying and establishing communication with data sources, collecting data in its native format, sending native format data to the Archive, converting data in native format to ARM standard format and sending to the Archive, reprocessing data when necessary, and ensuring data are of known quality and quantifiable uncertainty.

4.2 Development

The XDC development system is used primarily by XDC developers to design, write, and test software. The XDC development environment is modeled after the DMF development environment, and uses the same release procedures, with some additional software packages required for XDC data processing. XDC developers rely on a wide variety of utilities to develop and test forecast model, satellite, and surface data products. For example, there are a number of software packages available to read/write GRIB model data. Terascan software is needed to process satellite data provided by Seaspace.

All software run operationally begins on the development system. The software is tested in a pseudo-production mode on development by XDC operators to ensure that the software is working properly in an environment that mimics the production processing environment. Once testing is complete, the software is released and packaged for installation on the XDC production system. All software is under revision control and documented in Package Notification Forms (PNC) files, which include information such as input and output expectations, error messages, contact information and version information.

The development system also hosts the XDC software package repository, and a web-server for developing XDC quick-look visualizations. The development system can be configured to run the production processes if the primary production system fails.

5 XDC Data Flow Operations

External data are collected and processed to augment ARM data generated within the ARM Facility. This section describes the basic life cycle flow of external data through the XDC, beginning with identifying a data source to fill a ASR data need, setting up the data collection, processing, quality assurance, storage and archival, re-processing, and finally documentation.

5.1 Identifying Data for Science Needs

The scientific need and the priorities for acquiring, processing and archiving the external data streams are established by the science working groups and considers several criteria to determine the value added by bringing those external data into the ARM archive:

- Data are not archived anywhere else, or are difficult to retrieve.
- Acquisition cost
- The native data-format is difficult to use, so there is significant value added by converting to the ARM standard data format.
- The data are provided in huge datasets and extracting only the portion of interest to ARM improves the usability significantly
- Other value added by customizing the output data-products to ARM needs

Once a data set is identified as a product that is needed by the working groups, an Engineering Change Request (ECR) is submitted to start the development process. The XDC then works with the external data providers to coordinate data collection and to obtain any requirements and documentation needed for processing.

The XDC collects a wide variety of data types, including numerical weather prediction model output, satellite, and surface mesonet data, from a number of different external institutions. External data provide additional measurements and/or extend the spatial or temporal coverage

for ARM sites (Figure 2). Examples include the Rapid Refresh (RAP) numerical weather forecast model dataset, which is used in products to enhance and augment data collected from ARM sites. Spatial coverage is increased by many of the XDC datasets over the ARM sites, such as the global satellite data set from the Ozone Monitoring Instrument (OMI). Data products are also provided from a variety of data networks and organizations, which expand ARM's user community through data intercomparison and collaboration. Examples of organizations which provide data to the XDC include NASA AERONET, UCAR COSMIC, NCEP, ECWMF and the Oklahoma Mesonet.

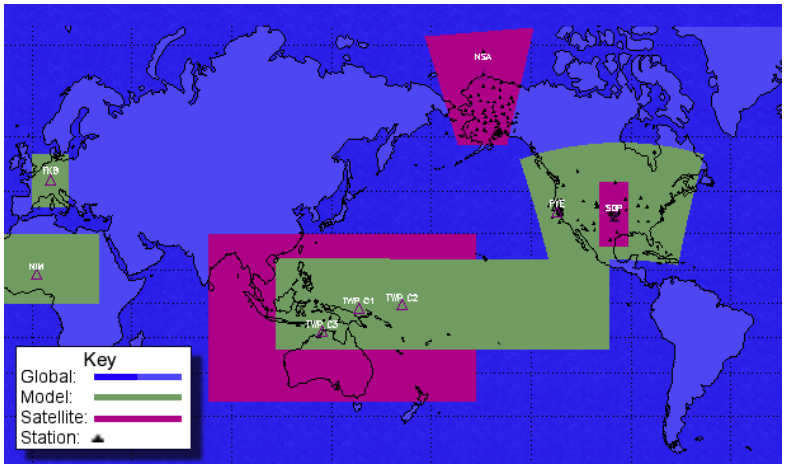


Figure 2. XDC data area coverage by data type

5.2 Collection

The XDC collects data directly from external data providers. The delivery method varies by provider. The XDC has evolved various collection methods to manage the diversity. The XDC determines the method based on ease and convenience for both the XDC and the external data provider. Most data are available on-line through ftp or the Web, and collection can be automated. The data collection flow is monitored via logs and tools such as DDtrack and DSview.

5.3 Processing

There are various kinds of processing performed at the XDC. Primarily, the XDC processes incoming external data to comply with ARM standards and to ensure the data are of known quality and quantifiable uncertainty. This includes converting data of different formats and various stages of pre-processing into ARM standardized NetCDF or HDF formats. Other types of processing include generating data images and running additional algorithms based on ARM scientists' requests.

Ingests are run daily, monthly or at irregular intervals throughout the year, according to data availability from the data source (e.g. high resolution quality controlled sounding data via NCAR.)

All ingests are controlled through automated scripts which send output to log files. The XDC operator reviews these log files for error messages and responds according to a given set of guidelines that accompany the ingest.

5.3.1 External data and related VAPs

External data sets differ from internal ARM data in that external data are often highly processed before arriving at the XDC. These data sets have gone through rigorous data transformations and quality assurance at their originating sources. Data that are not in ARM standard format are initially treated, however, as external "raw" data (level 00), though from a purely content perspective they could be equivalent to ARM's c-level value added products. When converted to netCDF or HDF they are assigned levels a1, b1, and c1 according to the content of the input data (see External Data data levels description in the ARM Standards Document).

In some cases, the XDC runs additional processes or creates Value Added Products (VAPs) to ensure data quality and to assist ASR scientists. For example, satellite composite images are created from GOES data to generate a quick view. A reduction algorithm is applied to NOAA's Rapid Refresh (RAP) weather forecast model data to reduce the data to a more manageable size.

Over time the XDC has incorporated a variety of off-the-shelf software programs to ingest raw external data, for ease-of-use or availability in the general user community. For example, GRADs software is used to convert European Center for Medium-Range Weather Forecasts (ECMWF) model data. This software is specialized to read/write the ECMWF GRIB data format and has the capability to convert data to netCDF. IDL is also used for ingesting data, because of its broad use in the scientific community, and IDL modules exist to interface with a wide variety of external data formats.

5.3.2 Reprocessing and Back-processing

Reprocessing external data is necessary when an update has occurred to the ingest, or more commonly, to the external data source (e.g. a modified algorithm, a new variable, or reanalysis after quality assurance or improved calibrations have been applied). Reprocessing efforts are coordinated with the ARM Problem Review Board (PRB), Reprocessing Advisory Board (RAB) and the ARM Archive. If the data streams are still available online at the XDC, processing takes place locally by XDC Operations staff on local XDC processing systems. If the reprocessing requires a large volume of input data which is no longer online at the XDC, reprocessing occurs at the Reprocessing Center by XDC staff. The Archive distributes a notification to any users that may have previously retrieved the data describing the reason for the update.

A new external data product most often begins the collection process before the ingest process is developed and tested. After the ingest is released it is typically applied to the newly arriving data operationally, and the previously collected data are "back-processed". This involves procedures identical to the reprocessing steps except no pre-existing data are being replaced.

5.4 Quality Assurance

Most external data have already gone through data quality processing prior to XDC collection at their originating sources. The XDC provides additional data quality monitoring, such as checking for missing data and mismatch dates before processing. The XDC submits data quality reports to track data problems and solution, as needed. In addition, the XDC also generates the composite images as online visualization of satellite data for additional quality assurance.

5.5 Storing and Archiving

Most of the data are stored at the XDC for 90 days. The forecast model and satellite data are stored for 30 days because of their large data volumes.

Data transfer from the XDC to the Archive and DMF occurs nightly. All data, from raw to the highest processed level are sent to the Archive via the Site Data System (SDS). Data needed as input to ARM VAPs are also sent to the DMF. Operators monitor the logs daily and review feedback from the Archive for data flow problems.

The Archive receives data from other sources including the DMF and the Data Reprocessing Center (DRC). It is the XDC's responsibility to coordinate data flow rates with the Archive. Data that are picked up at irregular intervals are sent to the Archive when ready via the STS.

In addition, the XDC also sends the Aerosol Observing System (AOS) QC Mentored data from BNL mentors to the DMF for the AOS VAP development when the data become available.

5.6 Documentation

To ensure the usefulness of ARM external data, the XDC staff strives to understand the data content and format and pass that information on to the end user. There are currently three levels of documentation.

- Notification forms containing information about data stream names, file sizes and temporal coverage are provided to the ARM Archive.
- Metadata information captured in appropriate databases
- More complete documentation is provided via the XDC documentation Web page located at https://www.arm.gov/xdc/xds/. This documentation is similar in format to the documentation provided for ARM Value Added Products (VAPs). However, some of the subheadings have been changed to better apply to external data. This documentation is useful both for ARM Infrastructure staff and to the end users of the data.

When all levels of documentation are completed the XDC announces the new data set to the ARM user community via the ARM Web page, in coordination with the ARM Publications Office.

5.7 General XDC Data Operations

The following are additional activities that support XDC Operations:

- Monitor daily operations via reports and tools:
    - Check the collection, delivery, ingest reports.
    - Check the instruments performance
    - Check the web and database servers performance.
    - Check the backup server and google server performance
    - Check the computer system load and disk space performance
    - Check DSview or DDTrack for data status.
- Notify the data clients, developers or system admin on problems of missing data, ingest errors or system performance errors.
- Submit DQPRs and DQRs for any issues discovered.

- Modify the existing data processing packages or create a collection or ingest for the new data, test the packages in the development system and install into production.
- Update the system software and hardware as needed.
- Provide as available weekend support to maintain 24/7 uptime. Keep an update emergency plan.
- Generate a monthly data transfer report.
- Monitor/coordinate BCR/ECR implementation impacts.
- Monitor and maintain the data systems according to ARM Cyber Security Plan

6 ARM Infrastructure Support

DMSE staff have been involved with designing DB applications since 1993, i.e. well before the XDC was conceived in 1996, and the scope has expanded as the project grew. This section describes the activities as of the beginning of FY 2015 in somewhat general terms. Specific technical details can be found on the ARM wiki (https://wiki.arm.gov/bin/view/Engineering/GeneralWebDev).

6.1 Database Support

The DMSE operates two dedicated database servers for ARM that currently provide production and development database engines running both Sybase Adaptive Server and MySQL. The databases include nightly copies of the ARM Integrated Database maintained at the ARM Archive, as well as locally developed and maintained databases supporting a number of ARM wide needs that can be categorized as “data quality”, “infrastructure” or “scientific”. They all have web-interfaces for administration of the database content and they typically drive dynamic web-content. The only database maintained in MySQL is the WordPress database that contains all the main ARM Web content including and especially the news items and facility/instrument descriptions.

6.1.1 Data Quality Databases

6.1.1.1 Data Quality Reports (DQR)

To document data quality issues, instrument mentors file Data Quality Reports that identify the specific data fields and time-ranges that are affected by the data quality problem. Processes in the data discovery tool access a copy of the DQR database to let prospective data-users know of any data quality issues at the time the user is looking for and requesting the data for research. With improved tools being developed to improve the ease and the accuracy of reporting specific time-ranges this database has the potential to grow to a large size.

6.1.1.2 Data Quality Assessments (DQA)

The staff at the Data Quality office review data for many instruments on a roughly weekly basis and capture their findings in this Data Quality Assessments database. These findings are shared with the Instrument Mentor and serve as reference to file DQRs or DQPRs.

6.1.1.3 Data Quality Problem Reports (DQPR)

When a data quality problem is first discovered it is entered into the DQPR database. Depending on the nature of the problem it is entered either as a "routine" DQPR, which then involves the specific site's operators and tasks them to try to diagnose and fix the issue, or, if the problem is beyond the scope of site operations and requires additional resources to address, the DQPR is "escalated" to the attention of the Problem Review Board, which then makes specific assignments to the staff that can diagnose and fix the issue.

6.1.2 Infrastructure Databases

There are several databases developed at the XDC that are primarily used by the ARM Infrastructure, though some are open to ARM users as well.

6.1.2.1 People database

PeopleDB stores personal contact information for all people that have interacted with ARM. It also provides information regarding email lists, group membership and roles within ARM. The account and group information is mapped to the LDAP service, which provides ARM-wide authentication and authorization (see Web Authentication and Authorization). The person ID in the PeopleDB is a primary key in many other database and web applications and is therefore a critical database.

6.1.2.2 Site Access Requests (SAR)

Anyone who needs access to an ARM facility or remote access to computer resources is required to submit a SAR. Since ARM became an official national user facility, it reports statistics about its users on a quarterly basis. The SAR administrator web application is used to retrieve this statistics from the SAR database.

6.1.2.3 Metadata Management

There are currently three databases and web-applications that are used to define site/facility, instrument class and category, and measurement metadata for ARM data-products: MMT for regular ARM data-products, IOP for Field Campaign data, and PIP for evaluation data-products, and principal investigator data.

6.1.3 Scientific Databases

The Poster, Publications, and Research Highlights databases support the reporting of the scientific results of ARM operations and ASR research and are cross-linked to both the ARM web-server and the ASR web-server.

6.1.3.1 Poster database

The Poster database stores information about the posters submitted to the ARM Science Team Meetings from 2007 to 2009 and, since 2010, for the ASR Science Team Meeting.

6.1.3.2 Research Highlights

ARM and ASR principal investigators are requested to submit brief summary reports about their recent research to this research highlights database.

6.1.3.3 Publications

The ultimate product of ARM operations and ASR research are journal and book publications. Authors that used ARM data or facilities are requested to share their publication citations with ARM. Those citations are tagged with ARM specific keywords identifying the ARM resources used and the research area and stored in the Publications database.

6.2 Web Server Support

The XDC installs, maintains and supports multiple web servers in support of the ARM program. There are two servers, one development and one production, running Apache software. Account setup, software installation and patching, system access, troubleshooting and technical support are all provided to web application and content developers on these two servers.

6.2.1 Web Site Services

The XDC hosts the following web sites as virtual hosts:

- Main ARM Web Site: www.arm.gov (www.dev.arm.gov)
- Main ASR Web Site: asr.science.energy.gov (dev.asr.arm.gov)
- DB Web Site: www.db.arm.gov (dev.www.db.arm.gov)
- XDC Web Site: www.xdc.arm.gov (dev.www.xdc.arm.gov)
- Education Web Site: education.arm.gov (dev.education.arm.gov)

6.2.2 Web Applications

A number of web applications provide critical forms for users and management tools for administrators.

User registration form and profile editor: A user may register as a guest or create an account. For active users with an account, the profile editor allows the user to manage their own personal profile.
PeopleDB admin tools: The PeopleDB administrator can manage user account information, email lists, and group memberships through the admin tools. There are also tools for group owners and email list owners to manage the group members and review the email lists.
Poster form: The poster form allows a user to submit new poster abstracts and full poster PDF files for ASR meetings, and an administrator to manage the submissions.
Publication form: The form allows a user to submit new publication entries and an administrator to approve and edit publication entries.
Research Highlight Form: The form allows a user to submit new research highlights and an administrator to approve and update research highlights.

6.2.3 Web Authentication

The XDC provides ARM-wide authentication through LDAP and Single Sign On (SSO). The main LDAP service is hosted at AmazonWS. There are couple of slave LDAP services located at BNL, DMF and others. The user accounts are mapped from PeopleDB and provide ARM-wide authentication.

The SSO service is deployed with Jasig Central Authentication Service (CAS). It is hosted at AmazonWS. It provides the advanced Enterprise Single Sign On Solution, and a overall better user experience because a user only needs to sign in one time for access to all SSO-enabled ARM applications.

Authorization is also achieved by using information about group membership maintained in the PeopleDB and LDAP service.

6.2.4 Web Content Management Database Administration

WordPress is used for the ARM, ASR, and Education web site content management. The XDC installs, maintains, monitors and backs up the MySQL servers. MySQL server is running as the backend supporting database for each web site. One production MySQL server and one development server is setup for WordPress and the Education web sites.

6.2.5 Web Search Engine

The XDC is responsible for identifying and maintaining a search engine for the ARM and ASR Web pages that is powerful enough to meet the search demands, yet cost effective. Currently the XDC provides a Google search engine, a Google Mini, on the ARM webservers for user searches over the ARM web pages. The XDC installs and configures the hardware and software, and monitors it daily to insure it is functioning properly. Periodic software and hardware upgrades are performed as necessary.

An index of documents is updated continuously, so the newest, and latest version of documents will be returned in the search results. On average, 25K documents are traversed in a 24-hour period. Documents are collected from www.arm.gov, www.db.arm.gov, iop.archive.arm.gov (Field Campaigns), BCR/ECR.arm.gov, and education.arm.gov.

User support is provided if needed.

6.3 Product Registration and Data Delivery

Principal Investigators (PIs) are required to register their products with the ARM Facility using the ARM Product Registration Tool (OME), developed by the ARM Archive, with standardized information about their data product. The types of products include field campaign data, PI products (e.g. ARM and ASR funded and externally funded research products) and value added products (VAPs). The XDC is managing the metadata process flow for each type of product, ensuring that corresponding databases are updated with the appropriate metadata, as well as communicating with Archive personnel who maintain the tool for improvements/corrections to the tool as necessary. These data are delivered to an unrestricted, public area of the Archive and organized and managed by XDC staff. The process is continually evolving as new DOE requirements for PIs are initiated. The tools and supporting databases are flexible and accommodating as new needs arise.

6.3.1 Field Campaign Data Products

The XDC manages the finalized field campaign data sent from investigators six months after the end of the field campaign. These data are delivered either through the OME tool interface, or if too large through an anonymous ftp server at the XDC. The XDC staff move the data to an unrestricted area. The OME xml form is converted to html and becomes the standard README file in the data directory. The metadata are assigned in the Field Campaign database so the data can be discovered through the Archive Data Discovery Tool and/or ARM Web pages, and so the end user understands the dataset and how it should be used.

6.3.2 PI Products

The ARM Climate Research Facility produces a number of datastream products to provide data of interest to the larger scientific community. These products are developed as part of the ARM baseline suite of instruments and value added products (see Section 6.3.3 below). Additionally there are scientists, both funded by ASR and those externally funded, who create products for their own needs that also prove beneficial to a wider group of scientists. If approved by the ARM Translators these datasets can become PI products, augmenting the ARM data products. The XDC verifies and manages the collection of the metadata for these products, updates the appropriate databases, and moves the datasets delivered via OME to an unrestricted area, allowing for data discovery and data understanding.

6.3.3 Value Added Products

ARM Value Added Products (VAPS) require OME registration of its metadata and data delivery as well, with a process similar to the previous product by the XDC. Each ARM VAP begins in an evaluation mode for a period of 6 - 12 months. During this time the metadata registered allow the evaluation product to be discovered from the ARM Web site where users can test the product and provide feedback. (See full discussion of VAP process in separate VAP Guidelines document.)

6.4 ARM Ingest and VAP Development

In addition to the ingest and VAP development for external data (as described in section 5 above), the XDC also develops ingests and VAPs utilizing the DMF's facilities. The XDC supports some of the ARM Mobile Facilities (AMF) Aerosol Observation System's (AOS) instrument ingest development. The AOS mentors are also located at BNL, easing communications between mentor and developer.

Since the AOS data are collected and delivered to the DMF, the XDC uses the existing DMF libraries and utilities to process the data into ARM standard format. In the event of any ingest problems, the DMF will contact the XDC for a resolution.

Currently, the XDC supports the following AOS instruments:

- HTDMA - Humidified Tandem Differential Mobility Analyzer
- ACSM - Aerosol Chemical Speciation Monitor, both raw and mentor-added data
- CAPS PMex - Cavity Attenuated Phase Shift Spectrometer-Particle Optical Extinction Monitor

6.5 ARM Site Surface Characterization

The XDC maintains surface characterization information in a geographic information system (GIS) database for the permanent ARM Climate Research Facility sites. The surface characterization database is available for ASR and the wider scientific community to request data or subsets of data for their research needs. These data include land use/land cover, soil texture, elevation, satellite and other remotely sensed data. For example, the surface characterization database has been used in the past for siting ARM Radars, based on such features as terrain and distance from the ARM Central Facility.

Additionally, a google map of all sites (current, past and future) is maintained on the ARM web page, including site layouts when available, lat/lon/altitude, and links back to ARM web pages. The Southern Great Plains site provides the highest level of detail, given its usage and the number of years it has been in operation.

6.6 ARM Metadata Management

Individual XDC and the Archive staff, named "metadata experts", manage the metadata process for the ARM Infrastructure. The Metadata assignment and approval process for ARM data is designed to review and improve metadata from the perspective of the scientific user community. This process is applied to metadata submissions for ARM data streams from routine ARM observations, field campaigns, PI data sets, and data sources external to ARM, as well as new sites and facilities. The process also provides evaluation and improvement of the metadata, including the continuing evolution of descriptive classifications for instruments and measurements, in particular instrument classes and primary measurement types. The goal is to provide the scientific user with consistent and concise pathways for data discovery when accessing the web-based descriptions of ARM data holdings.

Many of the ARM database applications and web applications, that organize the vast array of ARM metadata, were developed and are maintained at the XDC. It is beneficial that database/web application developers and a metadata expert are housed in the same location, for ease of communication.

7 Leadership Roles

Members of the XDC provide leadership and valuable "institutional memory" to the ARM Facility infrastructure organization. These include membership on the ARM Problem Review Board (PRB) , the Data Life Cycle Group (G6), the Engineering Review Board (ERB), and the ARM Standards Committee (Phase 2: Exception Review Committee).

8 Instrument Mentorship

The XDC is responsible for two instrument systems as ARM instrument mentors - the Cimel Sunphotometer (CSPHOT) and SuomiNet Global Positioning System (GPS). These instrument systems require close collaboration and coordination with external institutions for instrument deployment and data flow operations. The Cimel Sunphotometers (CSPHOTS) are deployed as part of both the ARM and NASA AERONET networks. The CSPHOTs are owned and operated by ARM. However, maintenance and calibration of the instruments are performed by AERONET. Most of the CSPHOT data are processed by AERONET before collected and ingested into ARM standards by the XDC.

The Suominet systems are deployed in collaboration with UCAR. The SGP GPS instruments are owned by ARM. The data are collected at SGP and NSA and sent for processing at UCAR before collected and archived by the XDC.

Instrument Mentorship responsibilities include coordinating maintenance, calibration, and deployment of instruments. Also, mentors review data quality on a daily basis, report all problems via the DQR and DQPR systems, and coordinate troubleshooting between ARM technicians, vendors and the external institutions (AERONET or UCAR) as required. Mentors maintain the instrument system documentation, including handbooks and metadata. In addition, mentors work with the science community to answer questions and support ASR science needs relating to the instrumentation.

9 Summary

The XDC manages the ARM external data data life cycle, including data source identification, data collection, processing, quality assurance, storage and archival, and documentation. In addition, the XDC operations staff provides database and web server support to ARM. The XDC manages the ARM Product Registration and Data Delivery process, including PI products, field campaign products and VAPs. Metadata Management activities are central to the XDC, as well as other leadership roles in ARM and instrument mentorship.

The XDC activities require a close and open relationship with many groups inside and outside of ARM to ensure they are accomplished smoothly and in a timely manner with data of known quality and quantifiable uncertainty.

10 Appendix A: XDC Roles and Responsibilities Table

DMSE/XDC Staff	Roles and responsibilities for Operations
Alice Cialella	XDC Manager, metadata coordination, Data Life Cycle Group, field campaign data, ARM surface characterization (GIS)
Bill Behrens	System Administrator, Backup Administrator
Laurie Gregory	Ingest/VAP data developer, CSPHOT and Suominet mentor
Kathy Lazar	Database applications, web applications, ARM Data Product Registration (OME)
Min Liang	Web services, web applications, database applications
Lynn Ma	XDC Operations Manager, database administration, external data operations, Suominet and CSPHOT mentor
Alison Tilp	Ingest/VAP data developer, field campaign data, ARM Search Engine
Richard Wagener	Metadata Management, PRB, ARM External Data identification, System Administrator, CSPHOT and Suominet mentor

11 Appendix B: XDC Mailing Lists Responsibilities

XDC Staff ARM Mailing Lists:

Mailing_List_Name	Description	Distribution
xdc_oper@arm.gov	XDC Operation	Alison, Lynn, Laurie, Rick
webmaster@arm.gov	Web Server Admin	Kathy, Laurie, Lynn, Min, Rick, Rolanda

webformadmin@arm.gov	Web Questions	Kathy, Rick, Lynn
google@arm.gov	Google Search Appliance Admin	Alison, Bill, Laurie, Rick

XDC and Additional ARM Staff Mailing Lists:

Mailing List	Description	Distribution
peopledb@arm.gov	People DB Admin	Alice, Kathy, Min, Lynn, Rick, others
peopledb_request@arm.gov	People Acct Admin	Kathy, Min, Jackie
www@arm.gov	Web Site Query Address	Kathy, Laurie, Min, Rick, others
asrwww@arm.gov	ASR Web Site Query	Rick, others
metadata@arm.gov	Metadata Notification	Alice, Kathy, Lynn, Rick, Archive
database@arm.gov	Database Problem	Bill, Alice, Kathy, Min, Lynn, Rick, others

12 Appendix C: Acronyms

AOS Aerosol Observing System
ARM Atmospheric Radiation Measurement
AMF ARM Mobile Facility
AERONET AErosol RObotic NEtwork
ASR Atmospheric System Research
BNL Brookhaven National Laboratory
BCR Baseline Change Request
COSMIC Constellation Observing System for Meteorology, Ionosphere, and Climate
CSPHOT Cimel Sun Photometer
DB DataBase
DDTrack Data Delivery Tracking
DMF Data Managemnt Facility
DMSE Data Management and Software Engineering group
DQR Data Quality Report
DQPR Data Quality Problem Report
DSView Data system Status Viewer
ECMWF European Centre for Medium Range Weather Forecasts
ECR Engineering Change Request
ERB Engineering Review Board
GrADS Grid Analysis and Display System
GOES Geostationary Operational Environmental Satellite
GRIB Gridded Binary
HDF Hierarchical Data Format

IDL	Interactive Data Language
IOP	Intensive Operation Period
LDAP	Lightweight Directory Access Protocol
MMT	Metadata Management Tool
OME	Online Metadata Editor
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
PI	Principal Investigator
PIP	Principal Investigator Products
PRB	Problem Review Board
SAR	Site Access Request
SSO	Single Sign On
UCAR	University Corporation for Atmospheric Research
VAP	Value-Added Product
XDC	EXternal Data Center