METEOROLOGICAL ASSIMILATION DATA INGEST SYSTEM TRANSITION PROJECT RISK REDUCTION ACTIVITY

James McNitt*, Joseph Facundo, and James O'Sullivan
National Oceanic and Atmospheric Administration's National Weather Service

7C.1

1. INTRODUCTION

The Meteorological Assimilation Data Ingest System (MADIS) is a data assimilation and distribution system developed by the National Oceanic and Atmospheric (NOAA) Office of Atmospheric Research Earth System Research Laboratory/Global Systems Division (ESRL/GSD) to perform the functions of data ingest, data integration, processing for QC, observation network monitoring, and data distribution for NOAA and non-NOAA observations worldwide. MADIS processes many types of meteorological data: surface observations from mesonets; upper air soundings, including balloon-borne and aircraft; remotely sensed data from both ground-based and satellite systems; and aircraft data. The MADIS development has reached a point where MADIS can now be transitioned to operations at the NOAA's National Weather Service (NWS) Telecommunications Operations Center (TOC) in Silver Spring, Maryland during fiscal year 2009. NOAA's NWS initiated its risk reduction activity in 2006. This paper describes the project risk reduction strategy and the results achieved through 2007.

2. MADIS OVERVIEW

MADIS ingests data files from NOAA data sources and non-NOAA data providers, decodes the data and then encodes all of the observational data into a common format with uniform observation units and time stamps.

Corresponding author address: James A. McNitt, W/OPS22, NOAA's National Weather Service, 1325 East West Highway, Silver Spring, MD 20910
Email: james.mcnitt@noaa.gov

Multiple quality control (QC) checks are conducted on the ingested data and the integrated data sets are created and then saved in MADIS. A series of flags are included in the data sets that describe the quality of the observational data. MADIS users and their applications can then inspect the flags and decide whether or not to use the observation.

MADIS provides the capability to visualize QC and station monitoring information through a graphical-user-interface called the Quality Control and Monitoring System (QCMS) Browser. The QCMS Browser is an application on Advanced Weather Interactive Processing System (AWIPS) workstations that provides easy access to the MADIS database and QC information for the purposes of monitoring station performance, locating persistent biases or failures in surface observations, and evaluating accuracy.

NOAA's ESRL/GSD has operated MADIS as a research system since 2001. In 2005, the NOAA Transition Board directed NOAA's OAR and National Weather Service (NWS) to transition MADIS from research to operations. The transition project and the resultant operational system are both referred to as MADIS Transition (MADIS-T).

3. CONCEPTUAL ARCHITECTURE

At its initial operating capability (IOC) the MADIS-T system will reside on the TOC floor space and interface to the National Weather Service Telecommunications Gateway (NWSTG) and the TOC's web services.

Figure 1 describes the general MADIS-T conceptual architecture. The architecture includes the existing MADIS functionality as the data integrator for a variety of diverse data feeds including NOAA and non-NOAA observation networks. MADIS-T conducts QC and integrates the various data sets and passes them on in the prescribed formats to NOAA customers through NWS telecommunications and web services.

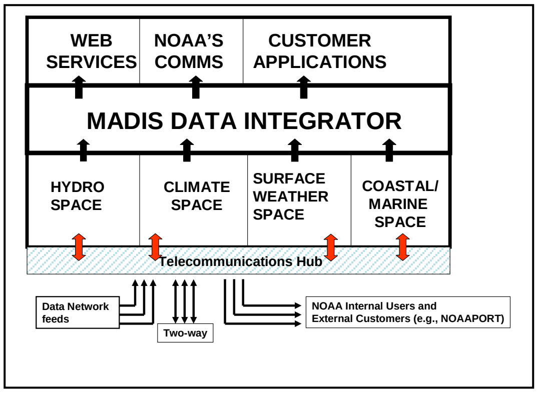

Figure 1. MADIS-T Conceptual Architecture

The architecture includes spaces for functions such as Fire Weather, Climate, Hydrology, Oceans, Surface weather, and Upper Air. Each functional space operates independently. For example, the climate functional space has its own set of applications to meet climate customer requirements and yet the data is integrated and disseminated to the wider NOAA customer base through MADIS-T. In this way, data are leveraged on different levels for different applications.

4. BLADE SERVERS AND VIRTUALIZATION

As described in the MADIS-T Concept of Operations and Operational Requirements Document the MADIS software is intended to be implemented in a systems architecture based on blade server and virtualization technology. Compute clusters based on blade servers will enable NOAA's NWS to control the MADIS-T operations and maintenance costs by providing a systems architecture that is highly flexible, cost effective, scalable, expandable, and reliable. Virtualization software is part of this target architecture

4.1 Blade Servers

A blade server is a server chassis housing multiple thin, modular electronic circuit boards, known as server blades. Each blade is a server in its own right, often dedicated to a single application. The blades are literally servers on a card, containing processors, memory, integrated network controllers, an optional fiber channel host bus adaptor and other input/output ports. Blade servers allow more processing power in less rack space, simplifying cabling and reducing power consumption. The advantage of blade servers comes not only from the consolidation benefits of housing several servers in a single chassis, but also from the consolidation of associated resources (like storage and networking equipment) into a smaller architecture that can be managed through a single interface.

4.2 Virtualization and Server Consolidation

Virtualization is a broad term that refers to the abstraction of computer resources. This includes making a single physical resource, such as a server, an operating system, an application, or storage device appear to function as multiple logical resources; or it can include making multiple physical resources, such as storage devices or servers appear as a single logical resource. Virtualization is performed on a given hardware platform by a host software which creates a simulated computer environment, a virtual machine for its guest software. By decoupling the physical hardware from the operating system, virtualization will enable NOAA's NWS to:

- Run multiple virtual machines with different operating systems at the same time on the same physical machine.
- Create fully configured isolated virtual machines with it's own set of virtual hardware to run an operating system and applications.
- Rapidly save, copy and provision virtual machines that can be moved from one physical server to another for workload consolidation and maintenance.

5. CONCEPT OF OPERATIONS

MADIS-T will consist of a Blade cluster connected to the NWSTG and other communications interfaces in the TOC. The MADIS-T system will provide secure access to system administrators to remotely access the system within the TOC environment. MADIS-T system will provide internal access to monitor the health of the blade system.

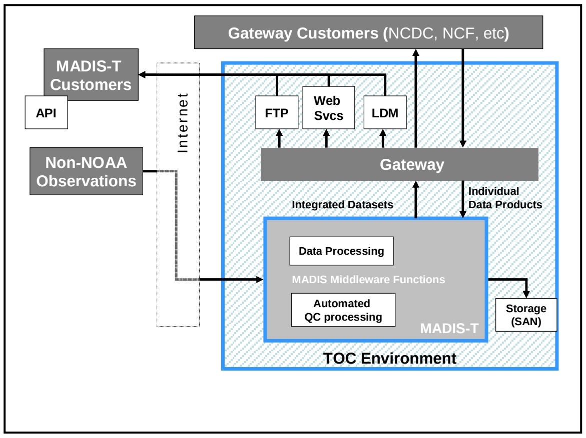

Figure 2. Notional MADIS-T Data Flow

The TOC will provide services to MADIS-T for data acquisition and distribution. Figure 2 describes the use of communication interfaces, front end processors, and the NWS Telecommunications Gateway (NWSTG) in the TOC without impacting the TOC operations. The NWSTG is the central data switching system for NOAA's NWS. These services could be provided within the current TOC environment with minimal or no impact on current TOC operations or could be implemented as part of MADIS-T.

There are two operational scenarios for data acquisition. In the first scenario the incoming data will either come directly to the NWSTG through the Internet or NOAANet. In the second scenario data sets would be received through a commercial host on the Internet.

Operational users will subscribe to MADIS-T web services or download MADIS-T data products from a data server, such as a Unidata Local Data Manager (LDM) server, in near-real time.

Data providers will download statistical data generated by MADIS-T to use for observation network monitoring. Statistical data can be used to locate persistent observational biases and hardware failures.

MADIS-T system monitoring could be incorporated in the TOC's monitoring capability. The help desk and system administration functions could be accomplished by an operations and maintenance (O&M) contractor provided by the system owner. Technology refresh would be an O&M function as well.

At IOC, ESRL/GSD will provide the backup capability for MADIS-T. At full operating capability (FOC), NOAA's NWS plans to maintain a backup MADIS-T at another location.

All changes will go through NOAA's NWS configuration management process prior to integration in the MADIS-T. Each change will be tested in both the development and pre-production environment before moving to production. All changes will be documented. Version control will be utilized to keep track of all MADIS-T software.

6. RISK REDUCTION ACTIVITY

The purpose of the MADIS-T risk reduction activity is to identify potential risk areas and conduct demonstrations to identify risk management strategies. The MADIS-T Core Design Team (CDT) is using the results of the MADIS-T risk reduction demonstrations to establish its technical approach, control costs, and develop work estimates. The MADIS-T CDT is an integrated work team comprised of experts from organizations within NOAA, other agencies, and contractors. It is responsible for designing and developing the MADIS-T system.

Risk reduction activity conducted in 2007 showed the benefits to NOAA of using modern server and virtualization technology to satisfy NOAA's real-time observational data processing requirements. Figure 3 lists the test cases that were demonstrated using a 16 node Hewlett Packard (HP) Linux cluster with virtualization software.

The MADIS-T CDT plans to use available high availability (HA) and failover solutions to increase MADIS-T reliability and availability. The HA software in MADIS-T will provide automatic failover from one node to another in case if application fails on a node. Virtualization will be used to host multiple operating systems on one node and it case of application failure, the complete node image will be moved to a spare blade. Similar techniques will also be used while performing maintenance.

All of the test cases in figure 3 were completed satisfactorily during the risk reduction demonstration. The demonstrated validated the NWS's conclusion that high availability software coupled with failover solutions will provide an end-to-end solution for increased system reliability.

Data Ingest & Processing

- *Hardware architecture*
- *Software architecture*
- *Data ingest*
- *Automated QC*
- *Product generation*

System Reliability

- *Application fail-over*
- *Storage fail-over*
- *Database fail-over*

System Security

- *Authentication and authorization*

Storage

- *Storage*

System Monitoring & Maintenance

- *System monitoring*

Management Software & Administration

- *System image provisioning*
- *Database support*
- *Database health and status*
- *General virtualization operations*
- *Physical to Virtual (P2V) functionality*
- *Hosting multiple operating systems with VMware Server*
- *Virtual drives for storage scalability*
- *Restoring VM states*

Web Services

- *Web server*

Figure 3. Test Cases for MADIS-T Risk Reduction Activity.

NOAA NWS's technical support team implemented and demonstrated a data processing thread in a blade cluster environment using virtualization software in mid-2007. Where the initial risk reduction activity focused on the hardware solution the second risk reduction activity is designed to show that the MADIS software can be hosted within the MADIS-T architecture as a data processing thread. Within the MADIS-T conceptual architecture, software on the blades will process multiple "data processing threads." A data processing thread represents the end-to-end processing of a specific dataset. The following functions are required to process a data processing thread: data ingest, data decoding, data processing, QC, temporary data storage, data monitoring, and data delivery. The satellite sounding and radiance thread described in figure 4 was successfully demonstrated in 2007.

7. FUTURE RISK REDUCTION ACTIVITIES

The MADIS-T CDT plans to use the results of future risk reduction activity to explore additional risk areas, such as IT security related risks. The results of IT security-related demonstrations will be incorporated into the MADIS-T design, implementation, and IT security plans. IT security planning includes the following: Risk Assessment

- Analysis of IT security functional Requirements (from FRD)
- Analysis of IT security system requirements (from SRS)
- Cost considerations
- Security planning
- Security control development during risk assessment
- Developmental Security Test and Evaluation (ST&E)

8. SUMMARY

Successful completion of the MADIS-T risk reduction activity provided an early demonstration of the benefits to NOAA and the Nation of transitioning the research capability developed by NOAA OAR to operations at NOAA's NWS in the TOC. The MADIS-T conceptual architecture includes data processing threads for surface, upper air, remotely sensed, and aircraft data. It also includes spaces for functions such as Fire Weather, Climate, Hydrology, Oceans, Surface weather, and Upper Air.

The evolving MADIS-T systems architecture is highly flexible, cost effective, scalable, expandable, and reliable. Blade servers and virtualization software are key parts of the systems architecture.

The MADIS-T IOC is planned for fiscal year 2009.

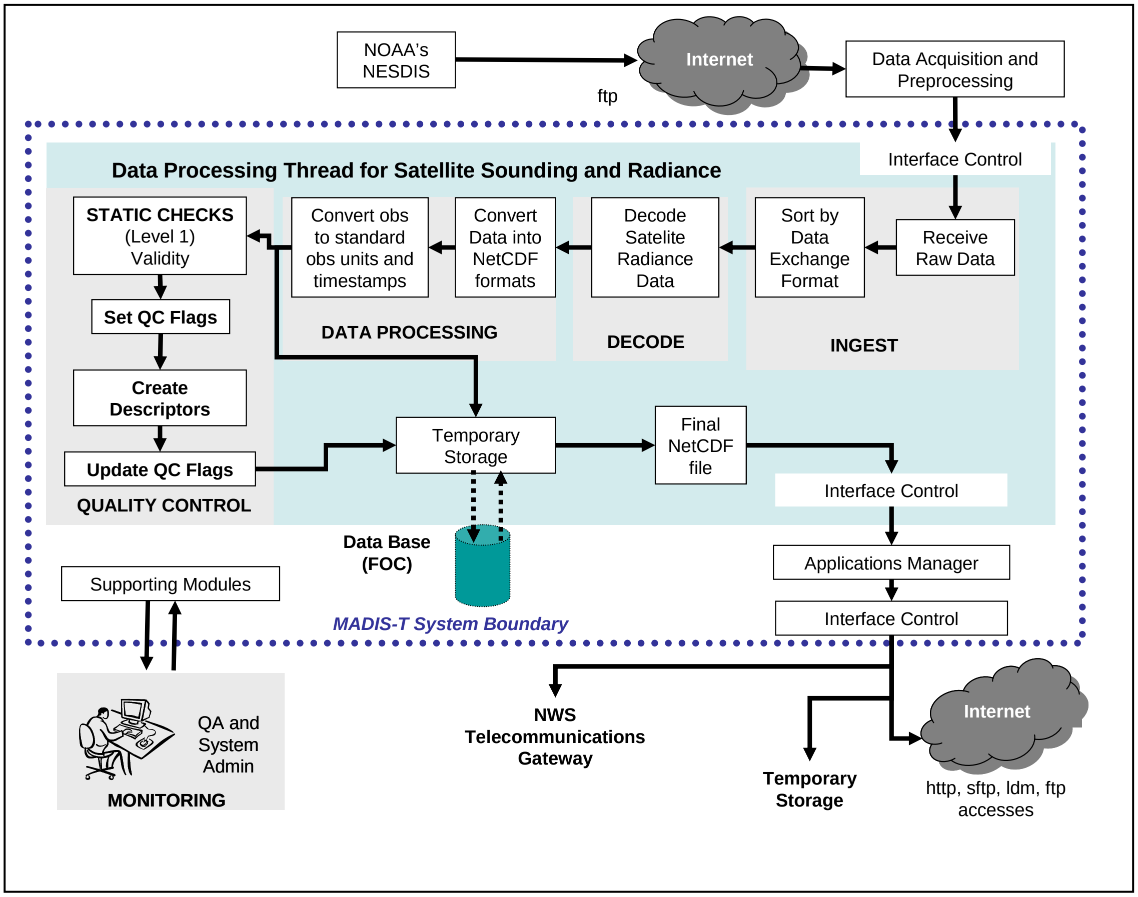

Figure 4. The Data Processing Thread for Satellite sounding and radiances.

9. REFERENCES

NOAA's National Weather Service, 2007: Meteorological Assimilation Data Ingest System Transition Concept of Operations and Operational Requirements Document.

NOAA's National Weather Service, 2007: Meteorological Assimilation Data Ingest System Functional Requirements Document.

NOAA's National Weather Service, 2007: Meteorological Assimilation Data Ingest System Transition System Requirements Specification.

NOAA's National Weather Service, 2007: Meteorological Assimilation Data Ingest System Transition Test Plan and Procedures.

10. ACKNOWLEDGEMENTS

The authors acknowledge the extraordinary success of the Earth System Research Laboratory Global Systems Division's MADIS and the positive impact MADIS has on MADIS users. Thanks to SAIC and Nortel Government Solutions for the successful risk reduction activity, Functional Requirements Document, and System Requirements Specification. The authors used text from those documents to compile this paper. The authors thank NOAA's NESDIS Office of Systems Development for the use of its equipment and acknowledge its groundbreaking work with blade servers.