CLARUS WEATHER SYSTEM DESIGN

SYSTEMS ENGINEERING ANALYSIS OF *CLARUS*-RELATED SYSTEMS

January 2006

Prepared By:

Mixon/Hill, Inc.
12980 Metcalf Ave, Suite 470
Overland Park, Kansas 66213
913-239-8400

Oklahoma Climatological Survey
100 East Boyd Street, Suite 1210
Norman, Oklahoma 73019
405-325-2541

Table of Contents

Table of Figures

Table of Tables

Revision History

Revision	Issue Date	Status	Authority	Comments
01.00	2005.09.26	Review	DTFH61-05-C-00022	Initial release for comment.
02.00	2006.01.27	Final	DTFH61-05-C-00022	Resolved comments.

Electronic File

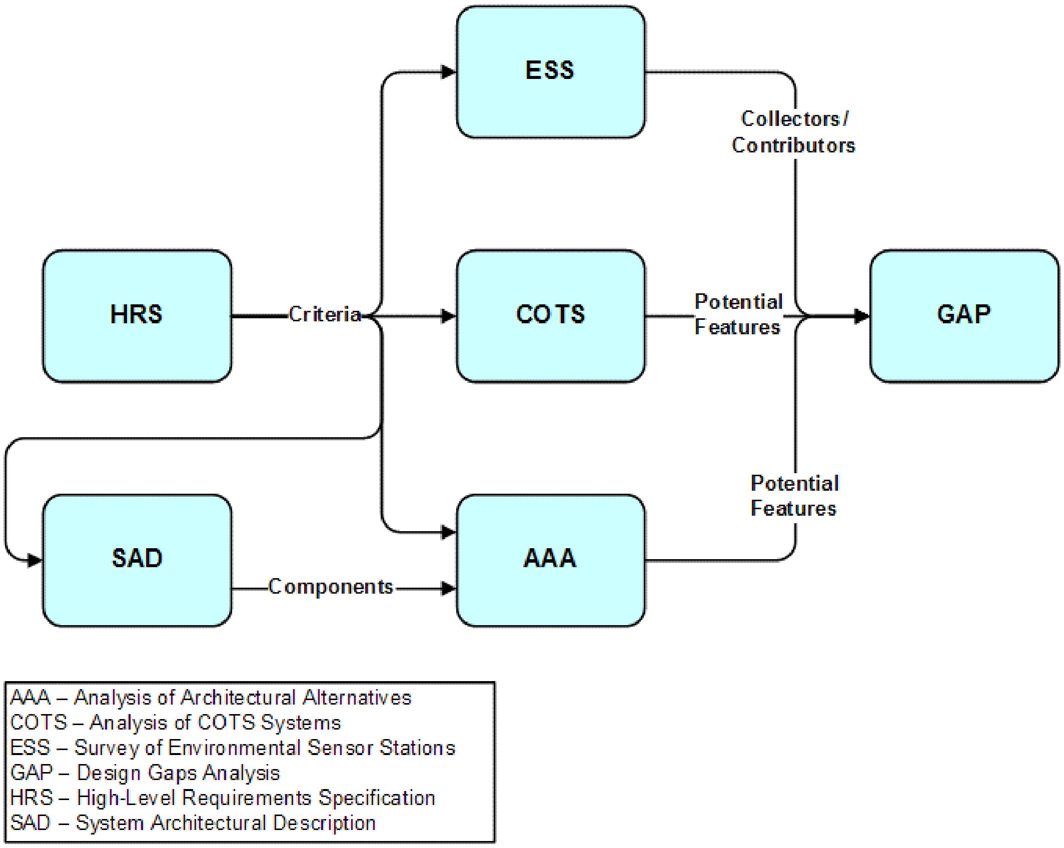
Saved As: 04037-rq301sea0200.doc

1 INTRODUCTION

1.1 Purpose

The purpose of this document is to provide an assessment of the relevance of existing (or “off-the-shelf”) software and hardware systems to the needs and design of the *Clarus* system. This assessment will contribute to an analysis of the gaps between existing surface transportation meteorological system capabilities and the requirements specified for the *Clarus* system. Existing systems (or components) that fulfill some part of the *Clarus* system requirements may be considered for inclusion in the *Clarus* system design.

This document is intended to be read and used by the U.S. Department of Transportation (USDOT) and system development team members. As indicated in Figure 1, the “COTS Analysis” is an intermediate deliverable in the larger context of the *Clarus* Weather System Design project, using criteria documented in the High-Level Requirements Specification to identify components and attributes (“potential features”) as input to the Design Gaps Analysis.



AAA – Analysis of Architectural Alternatives
COTS – Analysis of COTS Systems
ESS – Survey of Environmental Sensor Stations
GAP – Design Gaps Analysis
HRS – High-Level Requirements Specification
SAD – System Architectural Description

Figure 1 – COTS Analysis Context

1.2 Scope

This document provides an assessment of existing software and hardware systems with capabilities similar to or potentially contributing to the *Clarus* system objectives. *Clarus* will collect weather and pavement[1] condition information from environmental data sensors and mesonets. The system will qualify the environmental information using appropriate quality assessment methods to provide a relative indication of confidence in the information. *Clarus* will then format the qualified environmental information for dissemination to weather (or, more broadly speaking, environmental) service providers and for quality feedback to the environmental data contributors.

Clarus will provide benefits to a diverse group of stakeholders. Observing system owners and environmental equipment manufacturers will use *Clarus* information to improve the reliability and accuracy of their products. Transportation agencies will use qualified *Clarus* information to enhance their decision making in system operations and maintenance. Weather service providers, including the National Oceanic and Atmospheric Administration (NOAA), will use the qualified environmental information to enhance products distributed to the research community and the public.

1.3 Definitions, Acronyms, and Abbreviations

This document may contain terms, acronyms, and abbreviations that are unfamiliar to the reader. A glossary of these terms, acronyms, and abbreviations is provided in Appendix A.

1.4 References

1. *Clarus Final Draft Concept of Operations;* Iteris and Meridian Environmental Technology, Inc.; May 16, 2005.
2. *Clarus Weather System Design – High Level System Requirements Specification*; Mixon/Hill, Inc.; July 2005.
3. *Clarus Weather System Design – System Architectural Description*; Mixon/Hill, Inc.; August 2005.
4. *Clarus ESS Survey*; Cambridge Systematics, Inc.; September 2005.

Additional references are included in the discussion of each existing environmental observation network in Section 4.

1.5 Overview

This document provides an assessment of the relevance of existing software systems to the needs, architecture, and design of the *Clarus* system. User needs for the system, upon which the architecture and design are based, are documented

---

[1] "Pavement" in this context includes surface and subsurface data specified in NTCIP 1204 (Ref. 8).

in the Concept of Operations (ConOps) and further developed in the High-Level System Requirements. The high-level design is documented in the System Architectural Description.

The remainder of this document consists of the following sections and content:

Section 2 - System Description provides a description of the *Clarus* system.

Section 3 - Assessment Criteria describes the basis against which the various existing systems are to be assessed.

Section 4 - Existing Environmental Observation Networks provides an assessment of existing systems that perform functions similar to those specified for *Clarus*.

Section 5 - Commercial Environmental Data Collection Systems discusses some commercial offerings in environmental data collection that are prominent among state departments of transportation.

Section 6 - System Infrastructural Software discusses some of the server-level software components that may be needed to implement the eventual system design.

Section 7 - System Hardware discusses some of the hardware components that may be needed to implement the eventual system design.

Section 8 - Conclusions summarizes the findings of the assessment.

2 OPERATIONAL CONCEPT

The *Clarus* ConOps provides extensive discussions of the operational context, objectives, constraints, and system functions. These concepts are illustrated through discussion of an overall framework and operational scenarios for various user communities. Operational characteristics of the *Clarus* system itself are a subset of the overall framework and scenarios. The processes to be implemented in the *Clarus* system have been distilled from the framework in the ConOps and are shown in Figure 2 and Figure 3 below. This description focuses specifically on those functions to be fulfilled by the *Clarus* system and generalizes the interfaces based on the data types (rather than source types).

From the overall system perspective, the *Clarus* system will take in environmental data and metadata, and provide environmental metadata and qualified environmental data on request. The system will perform these operations based on data sharing agreements that define the terms of access and on quality control parameters used in assessing the incoming data. The system will need access to the environmental data networks and servers and will need to provide network access for users requesting information.

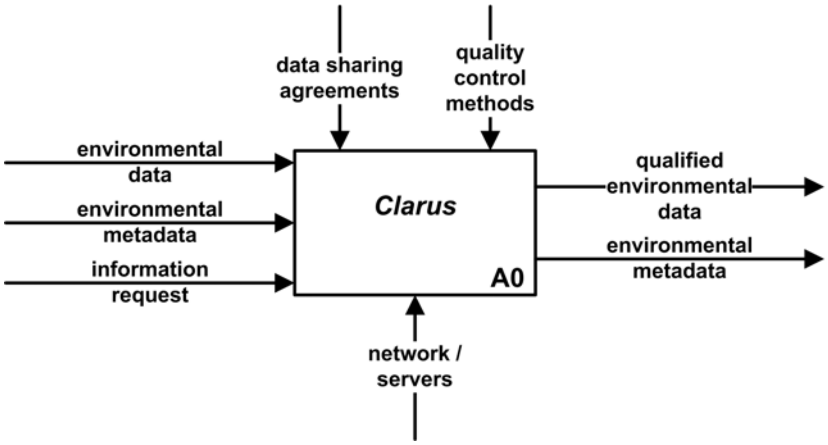


Figure 2 – *Clarus* System Process Context

Within its overall context, the *Clarus* system will collect, assess the quality of, and disseminate environmental data. The collection process locates, obtains, and stores the data in a common data structure, subject to access agreements. The quality control process applies one or more quality checks and associates quality flags with the data. The qualified data and the associated metadata are then available for dissemination, subject to any constraints specified in the data sharing agreements.

There will be multiple sources of data for the collection process, each potentially in its own format. Each source of data will also provide metadata describing the source and conditions surrounding the source. Terms under which data can be accessed from each source will be identified in data sharing agreements with the source organizations. Data are collected from the sources, interpreted from source formats, and stored in a common data structure.

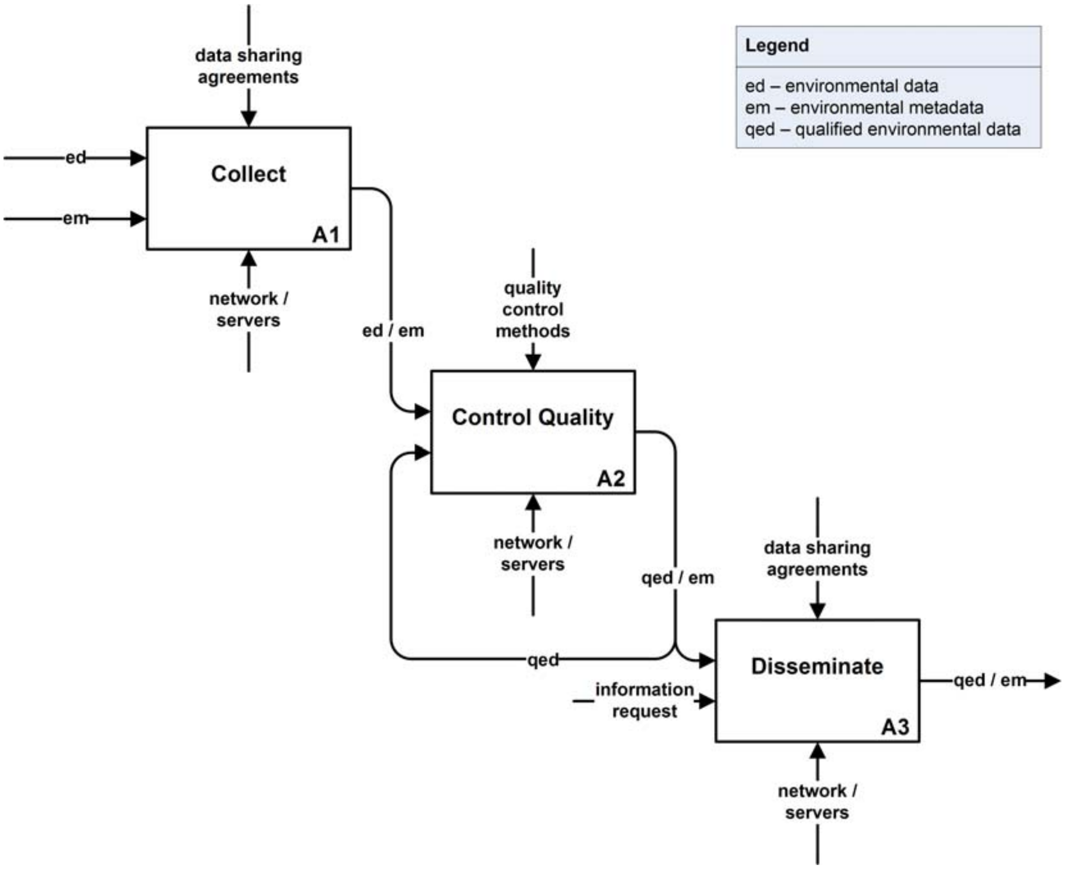


Figure 3 – High Level *Clarus* System Processes

The quality control process will implement one or more quality checks of the environmental data. Each quality check will be based on a set of rules for comparing the data to other models and data sets. Comparison data sets can include previously qualified data shown as a quality feedback loop in Figure 3. It may be necessary to derive or infer additional data from the original environmental data in order to complete some quality checks. Quality flags are assigned to the data according to the specific checks performed.

The data are disseminated in response to information requests directly from users with access to the data or automated processes based on data values and quality flags. In particular, data source organizations may be notified of data sets that do not pass particular quality checks. Data sharing agreements may constrain the data sets, formats, or distribution lists for data dissemination.

These essential *Clarus* functions provide the context for assessment of existing system capabilities relative to *Clarus* needs and system design.

3 ASSESSMENT CRITERIA

The value of any analysis depends on the care used in identifying criteria on which to base the assessment. Bad decisions are just as frequently made from inappropriate (or incomplete) criteria as they are from erroneous information. The body of criteria should consider the full context and scope of the assessment. In addition, care should be taken to assure that the criteria are based on needs, and not on abstracted descriptions of a solution. Objective criteria describe what the system needs to do, but do not prescribe what solutions might do it.

In this case, assessment criteria must address the *Clarus* operational concept, as captured in the high-level system requirements. These requirements provide a broad view of the context and scope of *Clarus* in terms of what *Clarus* will do, under what controls or constraints it will do it, and generally with what resources it will do it.

Requirements on the *Clarus* system contained in the High-Level Requirements Specification fall into four broad categories: functional, data, interface, and performance. Existing systems will be qualitatively assessed in these areas to gauge relative applicability to *Clarus* objectives.

- Functionally, the *Clarus* system will:
    - gather observations from fixed and mobile environmental data “collectors” across North America;
    - provide continuous quality control (i.e., assessment) and flagging of the data through a variety of methods, logging the quality control process;
    - disseminate data on request or by subscription, according to pre-established data sharing agreements; and
    - administer the process by managing access, issuing change notices, keeping logs, and retaining the environmental data according to the data sharing agreements.
- *Clarus* data will:
    - be based on the NTCIP, TMDD, and CMML standards;
    - include atmospheric data, pavement data, and hydrologic data;
    - include sensor metadata; and
    - include location, time and date stamp, and source for all observations.
- *Clarus* interfaces will:
    - be based on industry standards;
    - include provisions for direct manual data entry;
    - disseminate data based on a variety of query techniques, subject to data sharing agreements;

    - respond to dissemination requests on one-time and subscription bases; and
    - allow for system administration.
- *Clarus* will be designed to:
    - minimize latency in data collection, quality assessment, and dissemination;
    - scale to 470 million *current* observations;
    - scale to 600 concurrent users; and
    - scale to 6000 registered users.

As discussed earlier in this section, these criteria describe what the system will do, not how it might be designed to do it. The *Clarus* System Architectural Description documents perspectives on how the *Clarus* system could be designed to meet the requirements, but it is not itself a formalized design. To use the architecture as a basis for the assessment would overly prescribe an implementation. The architecture is therefore best used as a qualitative validation of the assessment, and not as a basis. If one of the systems being evaluated lines up well against the *Clarus* requirements, its architecture should look similar to the *Clarus* architecture.

4 EXISTING ENVIRONMENTAL OBSERVATION NETWORKS

Having defined a base for the assessment of the existing systems in Section 3, this section of the analysis provides the description and assessment of specific existing environmental data systems. The alternatives assessed herein include Federal agency-sponsored systems described in the *Clarus* ConOps (Section 3.6) and other environmental observation network systems sponsored by public agencies or universities. Table 1 summarizes the assessment by describing the systems' behaviors and attributes relative to *Clarus'* high-level requirements.

Table 1 – Selected Environmental Data Networks

[Y = Yes; N = No; N/A = Not Available]

	NETWORK	ACIS	ALERT / IFLOWS	ASOS	HADS	MADIS	Meso-West	NERON	NC ECONet	Oklahoma Mesonet	rWeather (Wash. DOT)	RWIN	USCRN	Weather-Share	West Texas Mesonet
Collects Observations from…	Collectors	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	N	Y
	Ground Vehicles	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	Weather service providers	Y	N	N	N	Y	Y	N	Y	N	N	Y	N	Y	N
	Across North America	N	N	N	Y	Y	N	N	N	N	N	N	N	N	N
Derived Variables	Calculates derived variables from obs	N	N	Y	N	Y	Y	Y	Y	Y	Y	N/A	Y	N	Y
Provides Quality Control	Continuously	N	N	N	N	Y	Y	Y	N	Y	N	N/A	Y	N/A	Y
	Multiple methods	Y	N	Y	N	Y	Y	Y	N/A	Y	N/A	Y	Y	Y	Y
	Logs the QC process	N/A	N	Y	N/A	Y	Y	Y	N/A	Y	N/A	N/A	Y	Y	Y
Disseminates Data…	On request	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
	By subscription	N	N/A	Y	Y	Y	Y	N	N/A	Some	N/A	N/A	N/A	Y	Y
	Based on data sharing agreements	Y	N/A	N	N	Y	Y	N	N	Y	N	Y	N	Y	N
System Administration	Maintains data sharing agreements	N/A	N/A	N	N	Y	Y	N	N	Y	N	Y	N	N/A	N
	Manages security groups	N/A	N/A	Y	Y	Y	N/A	Y	N	Y	N	N/A	Y	Y	N
	Issues system modification notices	Y	N/A	Y	Y	Y	N/A	Y	N/A	Y	N/A	N/A	Y	N/A	N/A
	Keeps operations logs and statistics	N/A	N/A	Y	Y	Y	Y	Y	N/A	Y	N/A	N/A	Y	N/A	Y
	Manages data retention (nominal 7 days)	Y	N/A	Y	Y	Y	Y	Y	Y	Y	N/A	Y	Y	Y	Y
Data Format Standards	Based on NTCIP 1204	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N
	Supplemented by CMML	N	N	N	N	N	N	N	N	N	N	Y	N	N	N
	Supplemented by TMDD	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Observation Types	Atmospheric	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Surface or subsurface	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Hydrologic	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

	NETWORK	ACIS	ALERT / IFLOWS	ASOS	HADS	MADIS	Meso-West	NERON	NC ECONet	Oklahoma Mesonet	rWeather (Wash. DOT)	RWIN	USCRN	Weather-Share	West Texas Mesonet
	Watches, warnings, alerts	N	Y	N	N	N	N	N	N	Y	N/A	N/A	N	Y	N/A
Metadata	Sensor	N	N/A	Y	N	N/A	Some	Y	N/A	Y	N/A	Y	Y	N/A	Y
	Geographic (location)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Time and date stamp for observations	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Source of observation	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Interfaces	Manual data entry	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
	Data dissemination	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Administration	N/A	N/A	Y	Y	Y	N/A	Y	N?A	Y	N/A	N/A	Y	Y	N/A
Data Dissemination	Select datasets of interest	Y	Some	Y	Y	Y	Y	Y	Y	Y	N	N/A	Y	Y	Y
	Query by timestamp	Y	N	N	N	Y	Y	Y	N	Y	N	N/A	Y	Y	Y
	Query by location	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N/A	Y	Y	Y
	Query by quality flag	N	N	N	N	N/A	N/A	N	N	Y	N	N/A	N	N	N/A
	Query by source	N/A	N/A	N	Y	Y	Y	N	N	N	N/A	N/A	N	Y	N
	Constrain by privilege (via data sharing agreement)	Y	N/A	N	N	Y	Y	Y	N	Y	N/A	Y	N	Y	N
	One-time request	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Subscription	N	N/A	Y	Y	Y	Y	N	N/A	Some	N/A	N/A	N/A	N/A	Y
Performance	Minimizes latency in data collection, QA, & dissemination	N	Y	Y	N	N	Some	Some	N	Y	Some	Some	N	Y	Y
	Operates with 470 million current obs	N	N	N	N/A	N	N	N	N	N	N	N/A	N	N	N
	Operates with 600 concurrent users	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Operates with 6000 registered users	N/A	N/A	Y	N/A	N/A	N/A	N/A	N/A	Y	N/A	N/A	N/A	N/A	N/A

4.1 ACIS (Applied Climate Information System)

4.1.1 Description

System Owner: NOAA’s Regional Climate Centers

Reference Site: http://rcc-acis.org/

NOAA’s Regional Climate Centers (RCCs) have developed an internet-based system to provide directed access for custom queries to climate data archives (Hubbard et al. 2004). The system, called the Applied Climate Information System (ACIS), is a distributed and synchronized system that provides redundancy of operations while ensuring timeliness of data availability. ACIS ingests, processes, and manages data in near real-time at the six RCCs. Data include NOAA’s Cooperative Observer Network, Hourly Surface Airways Network, and Historical Climatology Network as well as observations from non-NOAA mesonets (e.g., Automated Weather Data Network in the High Plains).

ACIS uses layers of independent modules that are interconnected by common object request broker architecture (CORBA). This design provides flexibility in the location and programming language of the modules (Hubbard et al. 2004). There are three interfaces designed for different levels of user sophistication and need. The low-level interface is CORBA; a mid-level interface is available in XML-RPC; and the high-level interface is web-based (HTML). All users must subscribe to ACIS, but they can receive the same data regardless of which RCC is used as an entry point.

Automated quality control uses four procedures that are tuned to the prevailing climate: seasonal thresholds (i.e., a range test), seasonal rate of change (i.e., a step test), seasonal persistence (Hubbard et al. 2005), and spatial regression. The spatial weighted regression test examines whether the given value lies within a confidence interval determined by the values at neighboring stations during a given time interval.

4.1.2 Analysis

ACIS results from a partnership between national and regional agencies and, hence, provides a nice example of a distributed system. ACIS’s strengths are its redundancy and synchronization across regional boundaries, use of multiple types of automated QC tests, modular design, and use of open standards. Because the system is focused on daily climate observations, the latency from measurement to data availability likely is too long for many real-time applications. ACIS is also a relative newcomer to the environmental data processing community and, as a result, it is difficult to assess the overall performance of the system. Except for its distributed nature, there currently are no unique features of this system to warrant close attention by *Clarus*.

4.1.3 References

Hubbard, K. G., A. T. DeGaetano, and K. D. Robbins, 2004: SERVICES: A Modern Applied Climate Information System. *Bull. Amer. Meteor. Soc.*, **85**, 811–812.

Hubbard, K. G., S. Goddard, W. D. Sorenson, N. Wells, and T. T. Osugi, 2005: Performance of quality assurance procedures for an Applied Climate Information System. *J. Oceanic Atmos. Tech.*, **22**, 105-112.

4.2 ALERT (Automated Local Evaluation in Real Time) and IFLOWS (Integrated Flood Observing and Warning System)

4.2.1 Description

System Owner (ALERT): NWS, USGS, Army Corps of Engineers, State and Local Governments, etc.

Reference Site (ALERT): http://www.alertsystems.org/

System Owner (IFLOWS): National Weather Service and partner organizations

Reference Site (IFLOWS): http://www.afws.net/

In the 1970s, the NWS California-Nevada River Forecast Center developed the Automated Local Evaluation in Real Time (ALERT) system to transmit environmental data from remote sensors to a central computer in real time. Today, ALERT systems are used by state and local governments (e.g., City of Denver, CO; Harris County, TX), the U.S. Army Corps of Engineers, U.S. Geological Survey, and U.S. Bureau of Reclamation, as well as the NWS. Typically, the remote stations include rain gauges, stream gauges, and other weather sensors. ALERT systems provide real-time data as a basis for flood warning systems.

The ALERT system provides a standardized protocol to transmit data using VHF radio from the remote station to a base computer (one-way communications). As a result, hardware and software vendors have developed interchangeable components that improve performance and reduce overall cost (Roark and Van Wie 2003). The remote stations can send data on a timed schedule or when a threshold event occurs (e.g., stream reaches flood stage). A personal computer receives the transmission with the aid of an antenna, receiver, and decoder connected to the serial port.

One example of processing software developed by the private sector is DIADvisor by DIAD Inc. (http://www.diad.com/). DIADvisor is a Windows-XP application that collects, displays, and stores ALERT and other real-time environmental data. Data are stored in Microsoft Access format for ease of creating reports. The host application uses FTP to transfer data to client machines networked via the internet, LAN, or WAN. Data also can be fed to the NWS's Local Data Acquisition and Dissemination (LDAD) system. The only automated check to the data is to verify that the data have been received during a given time interval. No automated routines check the validity of the data values themselves (OneRain 2005).

Another example is DataWise by DEC Data Systems (http://www.decdatasystems.com/). DataWise collects, processes, and analyzes environmental data from ALERT sensors using the 32-bit Windows environment with at least 128 Mb RAM and 100 Mb of disk space. DataWise can store data from up to 32,000 sensors in its custom database. Email alerts and map/graph displays can be output from the database using the DataWise software.

In the part of the eastern United States, the NWS supports the Integrated Flood Observing and Warning System (IFLOWS), based on the ALERT data protocol

(Gayl 1999). The architecture of the IFLOWS network is more complex than that of the basic ALERT system, but the ingest of the data into a base station computer, or control node, is similar to the ALERT system. Software that runs on the control node includes IFLOWS 4.6, a Windows 95/98-based application. Information on this software is not readily available.

4.2.2 Analysis

ALERT systems are widespread throughout the United States because of their low cost, real-time focus, and standardized protocols. The hardware has commoditized, albeit on 1970s-era technology. As a result, many implementations fail to take advantage of current communications and processor capabilities, although the ALERT community is positioned to make use of better communications. Further, the system offers no automated data verification routines. Therefore, there are no particular features of ALERT warranting close attention by *Clarus*.

4.2.3 References

Gayl, I. E., 1999: A new real-time weather monitoring and flood warning approach. Masters Thesis, University of Colorado, 82 pp.

OneRain, Inc., 2005: ALERT gaging system maintenance agreement 04-01.18. Annual report, December 28, 2004; revised January 20, 2005. Prepared for the Urban Drainage and Flood Control District, Denver, CO, 18 pp.

Roark, R. C., and D. Van Wie, 2003: The ALERT protocol: Evolving for the next millennium. White paper, 5 pp.

4.3 ASOS (Automated Surface Observing System)

4.3.1 Description

System Owner: National Weather Service, Federal Aviation Administration, and Department of Defense

Reference Site: http://www.nws.noaa.gov/asos/

The Automated Surface Observing Systems (ASOS) program is a joint effort of the National Weather Service (NWS), the Federal Aviation Administration (FAA), and the Department of Defense (DOD). ASOS serves as the nation's primary surface weather observing network and is designed to support weather forecast activities and aviation operations (NOAA et al. 1998).

Basic weather conditions measured by ASOS include sea-level pressure, altimeter setting, air temperature, dewpoint temperature, wind speed and direction, wind gust, precipitation accumulation, cloud height and amount, visibility, obstructions to vision (e.g., fog), and weather type and intensity for rain, snow, and freezing rain. In addition, alerts or special remarks are issued when conditions warrant, such as variable cloud height, variable visibility, precipitation beginning/ending times, rapid pressure changes, pressure change tendency, wind shift, peak wind. Additionally, ASOS routinely and automatically provides computer-generated voice observations directly to aircraft in the vicinity of airports, using FAA ground-to-air radio. These messages also are available via a telephone dial-in port. ASOS observes, formats, archives, and transmits observations automatically. ASOS transmits a special report when conditions exceed pre-selected weather element thresholds (e.g., the visibility decreases to less than 3 miles).

There are five types of ASOS data outlets:

1. On-site and remote interactive video screen displays;
2. On-site non-interactive screen displays;
3. On-site printer;
4. Coded messages; and
5. Computer-generated voice messages.

For users of the ASOS data, the coded and voice messages are the primary methods to receive output. ASOS observation messages are coded in METAR (Aviation Routine Weather Reports), SHEF (Standard Hydrometeorological Exchange Format), and daily/monthly summary formats. These coded messages are transmitted nationwide through NWS and FAA communications networks. Voice messages are primarily broadcast to pilots via ground-to-air radio. General aviation personnel also can access the voice messages by telephone.

ASOS quality control is completed in three levels:

1. On-site, automated;
2. Manually at the Forecast Office; and

3. Automated at the ASOS Operations and Monitoring Center (AOMC).

For the Level 1 QC, automated algorithms operate in self-diagnostic mode on the raw sensor data at the station. Should ASOS detect questionable data resulting from system degradation, component failure, or data error, the specific observation is omitted from the transmitted dataset and a “$” is appended to the transmitted report. After the ASOS data are transmitted publicly, personnel at the Forecast Office monitor and assess the availability and quality of the data from all ASOS sites within the office’s county warning area (CWA). This step occurs normally 1-2 hours after the observation was taken. The goal of this step is to reduce errors in future data transmissions and to inform maintenance personnel.

At the AOMC, Level 3 QC is conducted to identify the ASOS reports that contain “$” appended and those reports that are not received within their standard time interval. If a problem is encountered, an AOMC technician issues a trouble ticket, alerting the responsible maintenance technician. The AOMC is open 24 hours per day and provides a 1-800 phone number to report problems.

The National Center for Environmental Prediction (NCEP) generates automated QC messages hourly and delivers this information to the Forecast Offices. The QC reports result from an optimal interpolation (OI) of wind speed and direction, potential temperature, dewpoint temperature, and sea-level pressure (Miller and Benjamin 1992). Those elements that differ from the analyzed values (i.e., resulting from the OI) by more than a threshold amount are flagged as suspect.

ASOS data are available to the public through a host of conduits, including websites, NOAAPORT (satellite broadcast), and anonymous FTP (ftp://tgftp.nws.noaa.gov/data/observations/metar/decoded/).

4.3.2 Analysis

ASOS serves as the foundational surface observing system for the United States, operates on a substantial budget, and is monitored continuously by both automated routines and NWS Forecast Office personnel across the nation. ASOS’s strengths include its standardization of equipment and protocols, timely observations, systematic design, and continuity of operations. The ASOS program has not updated its automated quality assurance procedures for more than a decade, and non-NOAA customers have difficulty finding resolution to their reports of bad observations. ASOS appears to have no mechanism for reprocessing archived data to include manual QC flags or site maintenance information. Although ASOS provides the hourly data used by all weather service providers, there currently are no unique features of this network to warrant close attention by *Clarus*.

4.3.3 References

Miller, P. A., and S. G. Benjamin, 1992: A system for the hourly assimilation of surface observations in mountainous and flat terrain. *Mon. Wea. Rev.*, **120**, 2342-2359.

NOAA, DOD, FAA, and U.S. Navy, 1998: Automated Surface Observing System (ASOS) User’s Guide. National Oceanic and Atmospheric Administration, Department of Defense, Federal Aviation Administration, and United States Navy, 72 pp.

4.4 HADS (Hydrometeorological Automated Data System)

4.4.1 Description

System Owner: National Weather Service

Reference Site: http://www.nws.noaa.gov/oh/hads/

The Hydrometeorological Automated Data System (HADS) of the National Weather Service (NWS) exists to support the Flood and Flash Flood programs of the NWS Forecast Offices and the operations of the NWS River Forecast Centers (RFC). Partner agencies include the Water Resources Division of the U.S. Geological Survey, the U.S. Army Corps of Engineers, the Tennessee Valley Authority, the U.S. Bureau of Land Management, the U.S. Forest Service, the U.S. Bureau of Reclamation, and the States of California and Colorado.

HADS focuses on real-time and near real-time data acquisition, data processing, and data distribution (NWS 2002). HADS acquires raw hydrometeorological data from many of the 19,000 Geostationary Operational Environmental Satellites (GOES) Data Collection Platforms (DCPs). The DCP includes environmental sensors, datalogger, UHF transmitter, and a yagi antenna. DCP messages (e.g., environmental data) are up-linked to a GOES satellite and then down-linked to the GOES Data Collection System (DCS), maintained by NOAA's National Environmental Satellite, Data, and Information Service (NESDIS). From the DCS, the messages are sent via the NWS Telecommunications Gateway to HADS and other systems for wider distribution.

In most cases, DCPs send messages on an assigned channel every one to four hours that may include data from multiple times (e.g., 5-minute or 15-minute data). Some DCPs transmit observations every 5 minutes. For critical operations, such as alerts of high water levels, an unscheduled transmission of data in a short message may be completed on a different channel than the scheduled reports. The DCP uses precision GPS timers to control the clock at the station. Transmissions from the DCP last from 10 to 60 seconds.

The DCS at the NWS's Wallops Island Virginia Flight Facility receives DCP messages continuously. Raw GOES DCP data are available to HADS via the NWS Telecommunications Gateway within 3-4 minutes of the DCP transmission. Every 3 minutes, a HADS software component converts the raw messages into Standard Hydrometeorological Exchange Format (SHEF) products. SHEF encoded products are disseminated to NWS offices via AWIPS Satellite Broadcast Network and to external customers via NOAAPORT. HADS maintains a 72-hour archive of raw data. If there is a communications outage, backlogged data are retained until they arrive at HADS and other NWS data outlets.

Metadata used by HADS for each data collection site includes the latitude, longitude, DCP owner/operator, a DCP identifier, the types of variables measured and their units, the precision of the observed values, the time interval of observation, and the DCP uplink time and interval. This information is available to customers on a website.

Each NWS Forecast Office and RFC provides HADS a list of the observing sites and variables that HADS should transmit to their office. These lists are maintained in a relational database and can be updated by the local office via a password-protected web page. As a result, the content of the SHEF files is different from office to office. With the exception of flagging corrupt or incomplete transmitted records, HADS does not quality control or filter any data.

Perl and Java scripts interact with the HADS database to provide HADS users web access to site configuration information and both raw and decoded data. HADS maintains developmental, test, and production code separately and supports regular backup of database tables and software. System and software problems, including missing and suspect data, are reported to the Hydrologic Data Systems Branch of the Hydrologic Laboratory. HADS notifies its users of system and software upgrades in advance of deployment and upgrades are scheduled only from Monday through Thursday. HADS informational messages are sent daily via email to local NWS office contacts. These messages also are posted and logged on the HADS web page.

4.4.2 Analysis

HADS is the data conduit to receive hydrometeorological data through GOES data transmissions. The strengths of HADS include its disciplined operational environment, responsiveness to customers, and its post-processors for data formatting. HADS greatest limitation is that is does not provide automated quality control of observations. In addition, the system only obtains data from stations transmitting data through a GOES satellite — an expensive option for non-NOAA data providers. There currently are no features of HADS to warrant close attention by *Clarus*.

4.4.3 References

NWS, 2002: Hydometeorological Automated Data System Users' Information Handbook.

4.5 MADIS (Meteorological Assimilation Data Ingest System)

4.5.1 Description

System Owner: Forecast Systems Laboratory

Reference Site: http://www-sdd.fsl.noaa.gov/MADIS/

NOAA's Meteorological Assimilation Data Ingest System (MADIS) ingests, integrates, quality controls, and distributes data from a wide range of NOAA and non-NOAA hydrometeorological datasets. Since July 2001, MADIS has made these operational and experimental data publicly available to improve weather forecasting and to decrease the cost and time needed to transfer new data sets from research to operations. Data are obtained from about14,000 meteorological and hydrological surface stations, radiosondes, profilers, aircraft, radiometers, and satellites.

MADIS ingests data using an ASCII, comma-separated-value (CSV) format. Preprocessors translate data from their native formats to the required CSV files and observations are converted to standard units and time stamps. These files then are sent via LDM to an AWIPS Local Data Acquisition and Dissemination (LDAD) server for conversion to netCDF format. Required metadata include station name, variables measured and their units, latitude, longitude, and elevation. MADIS also collects other metadata, when available from the data provider. Weekly updates to station location information allows for sites to be added, moved, or removed from the processing system.

Static and dynamic quality checks of the data are specified primarily by NWS AWIPS Techniques Specification Package (TSP) 88-21-R2 (NWS AWIPS 1994). Static checks (i.e., single station, single time) are applied every 5 minutes to new data and include validity, vertical consistency, and internal consistency tests. Spatial, temporal, and position consistency tests comprise the dynamic, automated quality control (QC) procedures. The spatial consistency check (Miller and Benjamin 1992) executes every 15 minutes on new data and uses an optimal interpolation (OI) technique developed by Belousov et al. (1968). Humans can override automated QC flags using reject (for bad observations) and accept (for good observations) lists.

Data and quality flags are stored in a database, and netCDF output files are available through FTP, LDM (Unidata's Local Data Manager), OPeNDAP (OPen source project for Network Data Access Protocol), and web pages to about 350 users. These files are compatible with AWIPS display systems used at the NWS Forecast Offices. Because some data providers have restrictions on the redistribution of their data, MADIS provides three data distribution categories:

1. Access to NOAA organizations only;
2. Access to government, research, and educational institutions only; and
3. Access to all interested parties.

As a result of these categories, authentication is required for access by all users.

For users to easily access the observations and QC flags within the netCDF files, MADIS provides application programming interfaces (APIs), written in FORTRAN, for Linux, Windows, and several other Unix environments. Source code and compiled binaries are available at the MADIS web site. Users of the APIs can extract the wind data rotated to a specified grid projection, only read data that passes a specific QC level, or select data from specific networks or geographic regions (Barth et al. 2002).

Data processing occurs on 21 clustered, Intel-based machines that are running RedHat Enterprise Linux and are configured to have redundant backups. A failure in a primary system will result in redistribution of the processing to secondary computers. The processing nodes for automated QC are Dell 1750 computers with dual 3.2-GHz Xeon processors with 4 GB of RAM. These machines are attached to a Dell PowerVault 220 storage device with eight 15,000-RPM, 36-GB disks in a RAID 0+1 configuration (MacDermaid et al. 2005). Automated and human operator procedures monitor the status of incoming and outgoing data. Should incoming data not be available for 6 hours, an automated email is sent to the data provider. Operators also maintain data outage logs that are used internally. Computer and network resources are monitored using Nagios (http://www.nagios.org) and SolarWinds software (http://www.solarwinds.net), respectively. FSL personnel issue trouble tickets for computer failures and data output problems. MADIS generates hourly, daily, weekly, and monthly statistical reports on data failing the QC tests.

4.5.2 Analysis

Because MADIS currently provides RWIS and mesonet data to the road weather community, it serves as an excellent model for the *Clarus* system. MADIS's strengths include its data sharing agreements with a wide variety of agencies, its use of several different automated QC tests, its pre- and post-processors for data formatting, and the large number of stations available for forecasting and research applications. Performance measures are not well documented, but both experience with the system and user comments indicate that the latency from measurement to data availability is too long for many real-time applications. MADIS does not retrieve data directly from remote sites, and a significant amount of the data input in MADIS have been quality controlled prior to ingest. *Clarus* should continue to monitor the transfer of MADIS from its current research environment to the operational environment of the NWS.

4.5.3 References

Barth, M. F., P. A. Miller, and A. G. MacDonald, 2002: MADIS: The Meteorological Assimilation Data Ingest System. *Symp. on Observations, Data Assimilation, and Probabilistic Prediction*, Orlando, FL, Amer. Meteor. Soc., 20-25.

Belousov, S.L., L.S. Gandin, and S.A. Mashkovich, 1968: Computer processing of current meteorological data. Ed. V. Bugaev. Meteorological Translation No. 18, 1972, Atmospheric Environment Service, Downsview, Ontario, Canada, 227 pp.

FSL, 2004: An overview of the Forecast Systems Laboratory (FSL) Meteorological Assimilation Data Ingest System (MADIS). Draft document, Forecast Systems Laboratory, National Oceanic and Atmospheric Administration, August 12, 2004.

MacDermaid, C., R. C. Lipschutz, P. Hildreth, R. A. Ryan, A. B. Stanley, M. F. Barth, and P. A. Miller, 2005: Architecture of MADIS data processing and distribution at FSL. *21st Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, San Diego, CA, Amer. Meteor. Soc., CD-ROM, P2.39.

Miller, P. A., and S. G. Benjamin, 1992: A system for the hourly assimilation of surface observations in mountainous and flat terrain. *Mon. Wea. Rev.*, 120, 2342 - 2359.

Miller, P. A., M. F. Barth, and L. A. Benjamin, 2005: An update on MADIS observation ingest integration, quality control, and distribution capabilities. 21st International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology (IIPS), Amer. Meteor. Soc., San Diego, CA.

NWS AWIPS, 1994: Technique Specification Package 88-21-R2 for AWIPS-90 RFP Appendix G Requirements Numbers: Quality Control Incoming Data, 1994. AWIPS Document Number TSP-032-1992R2, NOAA, National Weather Service, Office of Systems Development.

4.6 MesoWest

4.6.1 Description

System Owner: The University of Utah and the National Weather Service

Reference Site: http://www.met.utah.edu/jhorel/html/mesonet/

Since 1994, researchers at the University of Utah and forecasters at the Salt Lake City NWS Forecast Office have collaborated to gather automated observations from both public and private sources throughout Utah (Stiff 1997; Slemmer 1998). These data are processed into a common format and redistributed to forecasters. In 2000, the program was expanded to include data from other western states (Horel et al. 2000; Splitt et al. 2001).

Currently, MesoWest gathers, processes, archives, integrates, and disseminates weather information from more than 2800 stations in the western United States. The average station spacing is about15 km in the combined network that comprises MesoWest. Data providers include the National Weather Service, U.S. Department of Agriculture National Resources Conservation Service, U.S. Forest Service, Bureau of Land Management, Bureau of Reclamation, state departments of transportation, several universities, wildland fire agencies, a host of local and statewide water agencies, and private companies.

According to Horel et al. (2002), the objectives of MesoWest are as follows:

- to improve timely access to real-time weather observations for NWS operations;
- to improve integration of observations for use in nowcasting, in forecast verification, and as input to operational and research forecast systems; and
- to provide access to available environmental data resources for research and education on weather processes in the western United States.

Depending on data availability from the networks, MesoWest servers query data providers for new observations from every 5 minutes to once per day using FTP, web retrieval, or LDM (Local Data Manager from Unidata). Synchronous data collection occurs every 15 minutes via FTP; asynchronous data collection also is used. The reporting interval typically ranges from 5 minutes to one hour for most data providers. The average data latency (i.e., interval between the valid time of the observation and the time the observation is available to users) ranges from 8 minutes for data from the NWS's Automated Surface Observing System, to 29 minutes for Remote Automated Weather Stations (RAWS), to 74 minutes for SNOWTEL (Snowpack Telemetry) sites. The stations from the remaining data providers have an average latency of 51 minutes.

MesoWest stores observations, metadata, and quality control flags in a MySQL database. Observations include air temperature, relative humidity, wind speed and direction, wind gust, precipitation (including snow depth and water equivalent), surface pressure, sea-level pressure, pressure tendency, altimeter setting, solar radiation, hours of sunlight, current weather conditions, visibility, cloud cover, cloud height and ceiling, soil temperature, soil moisture, lake temperature, fuel temperature, fuel moisture, and road temperature.

Station and sensor metadata are updated daily because some networks commission, move, or decommission stations constantly. At a minimum, station metadata include the station name, latitude, longitude, elevation, the variables measured, and their units. As of 2002, the database did not account for non-uniform metadata such as different mounting heights of wind sensors (Horel et al. 2002).

Automated quality control occurs on receipt of the data from the provider. Quality flags are issued based on range checks, simple algorithms, three-dimensional statistical regression (Splitt and Horel 1998), and a manual blacklist. A single quality flag ("good," "caution," or "suspect") is assigned to each observation (http://www.met.utah.edu/mesowest/quality/key.html) and the data then are labeled as "provisional." Provisional data are available free-of-charge via anonymous FTP (ftp://ftp.met.utah.edu/pub/mesonet/data) for governmental and educational use, including public safety, instruction, and research.

Maps, graphs, and tables are created from the data. Surface analyses incorporate both observations from MesoWest data providers and a background field from the 40-km RUC2 analysis of the National Centers for Environmental Prediction. MesoWest uses the Advanced Regional Prediction System Data Analysis System (ADAS) of the University of Oklahoma to produce 10-km resolution surface analyses every 15 minutes for the western U.S., 1-km resolution surface analyses every 15 minutes across northwest Utah, and 1-km analyses hourly over northwest Utah.

4.6.2 Analysis

MesoWest is a robust, collaborative network that obtains data from disparate networks and provides these data to a host of customers, including NOAA. MesoWest's strengths include its data sharing agreements with a wide variety of agencies, its use of several different automated QC tests, and the large number of stations available for forecasting and research applications. For several of the data sources, the latency from measurement to data availability is too long for many real-time applications. MesoWest also does not operate in a critical operations environment, so network/computer security and redundancy and automated system monitoring may not be adequate for around the clock operations. *Clarus* should continue to monitor the continued development of MesoWest.

4.6.3 References

Horel, J., M. Splitt, L. Dunn, J. Pechmann, B. White, C. Ciliberti, S. Lazarus, J. Slemmer, D. Zaff, and J. Burks, 2002: MesoWest: Cooperative mesonets in the western United States. *Bull. Amer. Meteor. Soc.*, **83**, 211-225.

Horel, J., M. Splitt, B. White, and L. Dunn, 2000: MesoWest—Cooperative mesonets in the western United States. *Preprints*, Ninth Conf. on Mountain Meteorology, Aspen, CO, Amer. Meteor. Soc., 265–268.

Slemmer, J. W., 1998: Characteristics of winter snowstorms near Salt Lake City as deduced from surface and radar observations. M.S. thesis, Dept. of Meteorology, University of Utah, 138 pp. [Available from Dept. of Meteorology, University of Utah, 145 South 1460 East Room 209, Salt Lake City, UT 84112-0110.]

Splitt, M. E., and J. D. Horel, 1998: Use of multivariate linear regression for meteorological data analysis and quality assessment in complex terrain. *Preprints*, 10th Symp. on Meteorlogical Observations and Instrumentation, Phoenix, AZ, Amer. Meteor. Soc., 359–362.

Splitt, M. E., J. D. Horel, and B. G. White, 2001: Data collection, processing, and quality control for MesoWest. *Automated Weather Stations for Application in Agriculture and Water Resources Management,* K. G. Hubbard and M. V. K. Sivakumar, Eds., WMO/TD No. 1074, 211–218 pp.

Stiff, C. J., 1997: The Utah Mesonet. M.S. thesis, Dept. of Meteorology, University of Utah, 138 pp. [Available from Dept. of Meteorology, University of Utah, 145 South 1460 East Room 209, Salt Lake City, UT 84112-0110.]

4.7 NERON (NOAA's Environmental Real-time Observation Network)

4.7.1 Description

System Owner: National Weather Service

Reference Site: http://www.isos.noaa.gov/

NOAA's Environmental Real-time Observation Network (NERON) is a new national program to install automated climate and weather observing stations every 400 square miles across the nation. Each station records air temperature and precipitation accumulation every five minutes and reports that information every 15 to 60 minutes. These stations are part of the National Weather Service's (NWS) effort to modernize the Cooperative Observer Program. This program consists of more than 11,000 volunteer observers who record daily readings of temperature and precipitation. These observations have formed the climate record of the U.S. but have not been available in real time. NERON is one facet of the NWS's Integrated Surface Observing Systems (ISOS) program.

Prototype operations and monitoring for NERON is conducted by the Oklahoma Climatological Survey (OCS) at the University of Oklahoma (McPherson et al. 2005). As a result, software developed for the Oklahoma Mesonet (*see separate description*) was modified for use by NERON. The primary differences between data ingest, processing, and dissemination for NERON and the Oklahoma Mesonet are as follows: (1) the number of measurements, (2) the dataloggers used, and (3) the data communications from remote site to centralized operations and monitoring.

One hundred stations have been installed in New England and eastern New York during the first phase of NERON. NERON requirements documents dictate that temperature and precipitation are the base measurements of these modernized, automated cooperative observer sites. Recently, wind speed and direction sensors have been installed on a number of the sites, and those observations have been incorporated into the ingest and processing system. Other measurements, such as soil temperature and soil moisture, may be added to the stations as the network evolves, but primary data ingest and processing is conducted on only a few variables. In addition, NERON uses a combination of Campbell Scientific and Vaisala dataloggers throughout the network. Nuances in data ingest and processing result from this heteorogeneous system (e.g., native data formats and some data logger algorithms differ).

The majority of the NERON observations are transmitted by satellite through the NWS's Hydrometeorological Automated Data System (HADS; *see separate description*). Currently, thirty-four stations communicate through a cellular phone connection whereby the remote sites are polled every 5 minutes by Campbell Scientific's LoggerNet on a 500-MHz Gateway Pentium 3 (with 384 Mb RAM) at OCS (Bain and Bostic 2005). Two sites connect directly to this LoggerNet computer via the NWS internet gateway. Two additional sites send data automatically every 15 minutes from Vaisala loggers to the Gateway computer via cellular phone. A Java applet written by Prism, Inc. runs every five minutes and combines the data from these 38 stations into a single file for ingest into the NERON ingest system. From this point forward, the system essentially is identical to that of the Oklahoma Mesonet, with differences attributed in XML configuration files.

Seven rack-mounted Dell Xeon 3.06-GHz processors (with 1 Gb RAM) running Linux are used for the processing environment (Bain and Bostic 2005). One computer runs ingest and automated quality control software; another generates output products. Two additional computers operate in the same configuration but in a testing environment. The final three computers are used as the database server, web/FTP server, and connection to the Apple Xserve storage device (running RAID 5 with extra disks as ‘hot spares’).

4.7.2 Analysis

Although primarily designed as the modernized cooperative observer network, NERON may become a national mesonet with more than 8,000 stations across the U.S. measuring data every 5 minutes. The strengths of NERON’s prototype operations and monitoring environment include its reliance on the well-tested software infrastructure of the Oklahoma Mesonet, its focus on obtaining quality observations, and built-in redundancies. However, the network currently operates only about 100 stations in the northeast U.S., it relies heavily on GOES transmissions with long data latencies, and its development is challenged by changing priorities and programmatic disagreements within NOAA. Because of NERON’s potential to become the nationwide mesonet standard in the next decade, *Clarus* should continue to monitor the development of NERON.

4.7.3 References

Bain, N., and Bostic, J., 2005: Personal communication.

McPherson R. A., J. M Wolfinbarger, and C. A. Fiebrich, 2005: Prototype a Modernized Data Ingest, Quality Assurance, and Monitoring System for NOAA’s Environmental Real-Time Observing Network (NERON). A proposal to the National Weather Service, Oklahoma Climatological Survey, 9 pp.

4.8 North Carolina ECONet (Environmental & Climate Observing Network)

4.8.1 Description

System Owner: North Carolina State University in partnership with state and federal agencies

Reference Site: http://www.nc-climate.ncsu.edu/econet/

The State Climate Office (SCO) of North Carolina, housed at North Carolina State University, has partnered with state and federal agencies to develop the North Carolina Environmental and Climate Observing Network (NC ECONet). When completed, the network will measure the near-surface environment at 100 stations across North Carolina. The NC ECONet is dedicated to serve the needs of North Carolina citizens and to assist in agriculture, emergency response, natural resource management, tourism, economic development, and education.

Currently, the SCO operates 26 stations that are supported financially by the NC Agricultural Research Service, NC Division of Air Quality, and NC Division of Emergency Management. In addition, the NC ECONet uses data from the National Weather Service, Federal Aviation Administration, and the U.S. Natural Resource Conservation Service.

NC ECONet stations maintained by the SCO send data via land-line at least once per hour. Data from the other agencies are obtained from the internet at least once per hour. The data are written to a MySQL database and can be retrieved via a web interface using PHP. Users can retrieve data based on station identifier, station name, county, or city. Data available include current observations, hourly data for the past seven days, and daily data for the past 30 days. Site metadata available include station name, station identifier, latitude, longitude, elevation, climate division, river basin, nearest city, county, site ownership, and date of first observation.

A Java-based tool allows for creation of web-based maps that display ECONet data. This tool requires the latest version of Sun Microsystem's Java Virtual Machine (JVM).

4.8.2 Analysis

North Carolina's ECONet is representative of many statewide efforts to develop mesonets on a very limited budget. These development efforts rarely are documented because personnel spend what little time they have trying to obtain as much data as possible for their customers. As a result, there are no performance metrics or system design specification that indicate any unique feature of NC ECONet that warrants attention by *Clarus*.

4.8.3 References

http://www.nc-climate.ncsu.edu/econet/

http://www.nc-climate.ncsu.edu/cronos/guide/

4.9 The Oklahoma Mesonet

4.9.1 Description

System Owner: The University of Oklahoma and Oklahoma State University

Reference Site: http://www.mesonet.org/

Since 1994, the Oklahoma Mesonet has operated more than 100 surface observing stations across all 77 counties of Oklahoma. The Oklahoma Mesonet is a joint program of the University of Oklahoma (OU) and Oklahoma State University, with primary operations located at the Oklahoma Climatological Survey (OCS) at OU (Brock et al. 1995). The Oklahoma Mesonet was established from the outset as a multi-purpose network with its primary focus to provide research-quality data in real time.

The Oklahoma Mesonet measures more than 20 environmental variables at 5, 15, or 30-minute intervals (depending on the variable). Data ingest, processing, quality control/assurance, and product development occur at OCS through an automated system that has been improved continuously over the past decade.

The Mesonet system polls each remote station every 5 minutes. Data are transferred from the remote site, via VHF radio at 4800 bps, to a nearby law enforcement agency. At each agency, connected to the Oklahoma Law Enforcement Telecommunications Network (OLETS), the data are transferred onto a TCP/IP stream running over a secure, encrypted Wide Area Network. These data then travel, via these OLETS connections and the OLETS Network (Switching) Center in Oklahoma City, to OCS for ingest into the Mesonet processing system. The minimum speed for OLETS' connections to its agencies is 56 Kbps, a sufficient speed for transmission of surface observations (as the daily-average bandwidth usage by the Oklahoma Mesonet is rarely over 60 bps). The Oklahoma Mesonet also includes two networks of 900-MHz, frequency-hopping spread-spectrum communications for locations where higher radio communications data rates are needed.

Data from Oklahoma Mesonet sites are ingested using multiple, redundant network connections to Campbell Scientific's LoggerNet™ monitoring software. The LoggerNet system employs four inexpensive x86 servers running Windows XP Professional™. Observation records are sent, via TCP/IP sockets, from LoggerNet to multiple instances of ingest and archival software running on four x86 Linux servers, where they are stored as both raw text and NetCDF archive files. Real-time and historical data ingest, processing, quality assurance, and product generation are distributed appropriately across these Linux servers. Because generic computers are used, any computer outage can be resolved simply by replacing the broken unit with an off-the-shelf spare. Data are archived on RAID units (currently Apple XServe RAIDs), utilizing RAID-5 sets and hot spare disks for redundancy and availability.

Oklahoma Mesonet data are never altered; each datum is flagged as 'good', 'suspect', 'warning', and 'failure' (Shafer et al. 2000). Automated flags are set using a three-stage process: (1) quality control (QC) filter, (2) QC independent algorithms, and (3) QC decision maker. The QC filter immediately flags data that fail the variable's range test, data that are known to be bad (as determined by a quality assurance (QA) meteorologist), and data coincident with a technician visit. The QC independent algorithms differ somewhat according to the variable observed, but include step, persistence, and spatial tests. In addition, like-instrument and variable-specific tests are applied when appropriate. The QC decision maker compiles the results of the independent tests and uses logic to determine the final automated QC flag assigned to the datum.

The automated QC software uses a single program for real-time, daily, and historical use through XML-based configuration files. For real-time data, up to eight QC tests are run per observation, operating on the past six hours of data. Once per day, up to 13 QC tests are run, operating on the past 35 days of data. As a result, more than 850,000 independent QC calculations occur every 5 minutes (for real-time data) and more than 111 million calculations are completed at the end of every day (for daily QC).

To manage both data and metadata, the Oklahoma Mesonet uses a MySQL™ relational database that is integrated throughout the Mesonet's system of data ingest, processing, and quality assurance (McPherson et al. 2006). The database has four interrelated components:

1. A user module;
2. A network site module;
3. An equipment module; and
4. A quality assurance module.

The database's user module is designed to store and retrieve basic information about each internal user type, including system administrators, lab technicians, field personnel, and quality assurance meteorologists. The network site module stores information regarding the overall network, details about each site (e.g., latitude, longitude), and both the types of variables measured and their attributes (e.g., sampling interval, units of measurement). Information about each instrument, radio, or other piece of network equipment is managed in the equipment module of the database. The calibration history includes the date of calibration, person who completed the work, and the resultant coefficients for every individual sensor.

The value of the previous database modules increases significantly when they are related to the quality assurance module. Should a data problem be detected, a QA meteorologist makes the final decision regarding how to "flag" the data and whether to send a technician to fix a problem. If a physical repair were necessary, the QA meteorologist issues a "trouble ticket" that outlines the problem, the date the problem was determined, the latest date when a repair must be made, and the data quality flag to be assigned until the problem is resolved. The trouble ticket is sent automatically to the technicians, who then can retrieve all previous database entries for the specific sensor and can submit repair information (e.g., date/time fixed, diagnosed error, resolution of problem). The information submitted by the technician becomes available to the quality assurance meteorologist so that is can be included in the database and concluding analysis can be performed of the equipment problem.

Data dissemination primarily occurs through password-protected FTP and both public and password-protected web pages. Real-time and archived data are available freely for non-profit use by Oklahomans and for public safety use by external organizations. To provide products in as near real-time as possible, the Oklahoma Mesonet limits the file distribution to only a few formats, primarily ASCII tables or XML-based files. Gridded products, such as the Oklahoma Fire Danger Model, are distributed in GIF image format. Customizable display of Oklahoma Mesonet data, other data sources (e.g., radar, lightning, ASOS, satellite), and geographical information (e.g., highway shapefiles) is provided through the WxScope Plugin (a web browser plugin) or WeatherScope (a stand-alone application). These display tools are available for Windows and Macintosh operating systems (http://sdg.ocs.ou.edu/).

In general, data arrive at the Mesonet's central computing facility 3 to 3.5 minutes after the observation was measured in the field. Within about1 minute, the computer system ingests and quality controls the data; within 1 to 1.5 minutes, it generates products, from least complex (e.g., ASCII data files) to most complex (e.g., fire danger model output). More than 65,000 products are generated daily and served from three load-balanced Apache web servers. More than 150 Gb of Mesonet data files were served to customers during 2004. These users downloaded an additional Terabyte of related data and products (e.g., agricultural model output, radar data) in the same time frame.

Software modules used internally by OCS personnel are written primarily in C++, using Open Source libraries such as OpenGL. Scripts that link data to the web pages primarily use Perl and PHP.

4.9.2 Analysis

The Oklahoma Mesonet is funded at a level greater than other state mesonets and, as a result, can operate in a highly disciplined, around the clock environment with standardized equipment. Strengths of the Oklahoma Mesonet include its combined automated and manual quality assurance procedures, large and diverse customer base, modular software design, and minimal latency from measurement to availability of quality assured data. Because the Oklahoma Mesonet is an end-to-end system, it can control the quality of the instruments in the field and, as a result, does not have experience with quality control of data from poorly maintained sites. In addition, the Oklahoma Mesonet has no experience ingesting or quality assuring data from road pavement sensors. Only a few types of data sharing agreements have been implemented by the Oklahoma Mesonet. *Clarus* should continue to monitor the development of the Oklahoma Mesonet.

4.9.3 References

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The Oklahoma Mesonet: a technical overview. *J. Atmos. Oceanic Technol.*, **12**, 5-19.

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonet, *J. Atmos. Oceanic Technol.*, **17**, 474-494.

4.10 rWeather

4.10.1 Description

System Owner: Washington Department of Transportation, the University of Washington, the National Weather Service, and the Northwest Weather Consortium

Reference Site: http://www.wsdot.wa.gov/traffic/

In 1998, the Washington Department of Transportation (WSDOT) initiated rWeather, a program to expand and improve weather data available to both WSDOT maintenance personnel and the traveling public (Boon and Cluett 2002). Currently, rWeather includes about60 Road Weather Information Systems (RWIS) stations owned and maintained by WSDOT and almost 400 other stations owned by nine other federal, state, and local agencies.

The components of the RWIS include sensor stations along roadsides, a statewide communication network, data access tools, weather forecasting and modeling services, pavement temperature modeling and prediction, and an Internet website for maintenance decision making and traveler information (Boon and Cluett 2002). All stations measure air temperature, wind speed, and wind direction; many of the stations also measure road surface and subsurface temperatures; and some stations detect the presence and concentration of snow- and ice-control chemicals on the road surface. The latest version of the website employs Java and ColdFusion technologies for display of data (WSTC 2002).

Data are sent from the remote station to a central computer using a remote data collection and processing unit (RPU) from Surface Systems, Inc. (SSI). The RPU supports direct wire, phone line, radio, radio/microwave, spread spectrum, fiber optics, and CDPD communications. SSI products use open source technologies and are NTCIP-ESS compliant.

Quality control tests for meteorological observations are implemented by the University of Washington (http://www.atmos.washington.edu/mm5rt/verify.html) and were modeled after those described in Shafer et al. (2000). These tests include range, step, persistence, and spatial checks. Daily quality control logs are available about 3 days after the date of the observation.

4.10.2 Analysis

The rWeather system provides important and timely data to DOT maintenance managers and the traveling public of Washington. It represents similar RWIS networks in other states, though its post-processing use of quality assurance tests at the University of Washington is different from most state RWIS networks. Unfortunately, as with most statewide RWIS networks, the data processing system is proprietary and, as a result, it is difficult to assess system performance and scalability. There are no unique features of rWeather to warrant close attention by *Clarus*.

4.10.3 References

Boon, C. B., and C. Cluett, 2002: Road Weather Information Systems: Enabling Proactive Maintenance Practices in Washington State. Research Report, Research Project T1803, Task 39, Road Weather Info, Washington State Transportation Center, University of Washington, 126 pp.

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, 17, 474-494.

Washington State Transportation Center (WSTC), 2002: Redesigned rWeather Site Offers More Information, Easier Use. *rWeather Newsletter*, Vol. 6, Winter 2002, 10 pp.

4.11 RWIN (Road Weather Information Network)

4.11.1 Description

System Owner: Meteorological Service of Canada

Reference Site: http://www.msc-smc.ec.gc.ca/

Environment Canada's Road Weather Information Network (RWIN) is a Canada-wide integrated network being designed and implemented to obtain, quality assure, and disseminate data from Environmental Sensor Stations (ESS) and instrumented vehicles. The purpose of RWIN is to provide a more efficient, sustainable, and safer highway system for Canada (Koonar et al. 2004).

The Provincial Computer Center ingests environmental data from fixed and mobile platforms and then transmits ESS data (fixed platforms) to a central RWIN system. RWIN assures the quality of the data, integrates data from heterogeneous sources, stores the data in a long-term archive, checks the health of the national ESS network, and reports outages and data quality problems to regional maintenance staff.

RWIN is designed as a distributed system using client/server applications and component-based architecture. The architecture is scalable and capable of including text, digital images and video, GIS layers, and raster data sets. Web dissemination supports spatial data browsing, map displays, and both interactive and batch mode data exchange. Object-oriented, application software is being developed using J2EE and the Java2 programming language. There are six subsystems designed to support the functional requirements of RWIN:

1. Data acquisition and processing (DAP);
2. Quality control (QC);
3. Alert manager sub-system;
4. Data manager sub-system;
5. Data exchange sub-system (DES); and
6. Maintenance sub-system.

The DAP sub-system receives data from partners in Canadian Meteorological Markup Language (CMML) format, converts any non-CMML data to a common format, and inserts the observations into a database. In addition, the DAP translates all imagery into a common map projection and converts model output to a specified grid spacing. DAP computers are clustered and load balanced to complete operations according to strict deadlines.

The QC sub-system checks the format of the incoming data files, scans for missing observations and data elements, performs range and temporal tests, and conducts inter-sensor checks on both atmospheric and pavement data. Possible QC flag values include accepted, missing, error, doubtful, inconsistency, and QC test not performed. QC of pavement temperature data will compare observations to values modeled from atmospheric data, compare pavement and air temperature observations, and check the range based on climatology (Delannoy and Koonar 2005). The results of the QC checks result in a quality-controlled dataset and a set of rejected data, both of which are added to the data repository, separate from the raw data.

The alert manager sub-system generates internal and external notifications of events, based on specific criteria. Warnings or error messages can be delivered through the DES via email, paging, or internet broadcast. Errors range from data problems to computer outages. The alert manager includes libraries of alert codes and personnel contact profiles. The severity of the problem is evaluated dynamically and the alerts are issued accordingly.

The Oracle™ 9i data repository, encapsulated by the Enterprise Data Interface (EDI), includes a relational database management system (RDMS) and a metadata catalogue. The RDMS provides automatic indexing, low-level data security, and storage of instrument observations. Image data are saved to files rather than the database. Data objects are accessed using a Uniform Resource Locator (URL) that defines a unique address for any data object and the required access protocol. Site metadata will include latitude, longitude, elevation, physical description, soil type, sensor type and make, construction of the roadway, contact information, photographs, and dates of commissioning, maintenance, inspections, and instrument calibrations. All RWIN metadata are stored as geospatial, being accessed by spatial coordinates (e.g., latitude, longitude), using the industry-standard ESRI shapefile specification. The EDI provides reusable, data-level software components to create more sophisticated applications. This interface provides constraints on the addition, modification, access, and deletion of data in the repository.

The DES packages information according to the need of the user and distributes the product via FTP, email, paging, or web sites. Both interactive and batch modes are available, and users can specify the data region, types of observation, and data format. ESRI's Internet Management System (ArcIMS) handles the user-driven data requests. The DES employs open standards such as Extensible Mark-up Language (XML) for data structure, Graphical Mark-up Language (GML) for data display, Simple Object Access Protocol (SOAP) for data transfer, and Web Service Description Language (WSDL) and Universal Description Discovery and Integration (UDDI) for availability of data services. Data distribution is governed by data sharing agreements with the agencies that collect observations.

Finally, the maintenance sub-system serves as the administrator of all of the other RWIN subsystems. The interfaces in this sub-system will allow personnel to manage and monitor all software and hardware components of RWIN, view station logs and history from any ESS, manage user and administrative profiles, and generate performance reports.

RWIN takes advantage of a host of accepted, proven standards to increase flexibility and ease-of-use and to decrease development and maintenance time. The following standards have been adopted by RWIN:

- CMML, an XML schema for encoding and transferring ESS data;

- Geographic Mark-up Language (GML), an XML schema for encoding geospatial data;
- National Transportation Communications for ITS Protocol (NTCIP), a set of standards for the communication of data to and from roadside devices;
- ISO Geographic Metadata Standard (ISO-19115:2003), defining the schema for describing geographic information and services;
- Scalable Vector Graphics (SVG), an XML-based standard to draw graphics via the web; and
- Unified Modeling Language (UML), a standard for creating diagrams used in system development.

4.11.2 Analysis

RWIN is an excellent model for *Clarus*, as the program is developing similar capabilities across a nation with diverse climate regimes and harsh environments. The best attributes of RWIN are its use of open standards, its focus on quality assurance of road pavement data (in addition to weather data), and its modular development. Unfortunately, RWIN has not been tested operationally across Canada as of this writing, nor have all of the functionalities of the system been developed yet. As a result, it is difficult to assess the performance of the system, especially for obtaining quality assured data to the users very quickly after the measurements are taken. *Clarus* should continue to monitor the development of RWIN.

4.11.3 References

Delannoy, P., and A. Koonar, 2005: Road Weather Information Network (RWIN) Update. A presentation to the *Clarus* Interagency Coordinating Committee (ICC), March 2, 2005, Las Vegas, NV, 18 slides.

Koonar, A., P. Delannoy, and D. Denault: Building a Road Weather Information Network for integrating data from heterogeneous sources.

Koonar, A., P. Delannoy, and B. Scarlett, 2004: A real-time Road Weather Information Network (RWIN) Spatial Data Management and Information Delivery architecture. GML Days – July 2004, 9 pp.

4.12 USCRN (U.S. Climate Reference Network)

4.12.1 Description

System Owner: National Climatic Data Center

Reference Site: http://www.ncdc.noaa.gov/oa/climate/uscrn/

The United States Climate Reference Network (USCRN) is a network of climate observing stations under deployment as part of a NOAA research initiative (McGuirk 2003). The purpose of USCRN is to provide and maintain long-term (50-100 year), high-quality observations of temperature and precipitation to meet the stringent data quality and continuity requirements of the climate science community.

This network of about 110 high-quality climate-observing systems has been located nationwide to observe temperature and precipitation across the climate regions of the U.S. The initial system is designed to accommodate a possible expanded array of sensors such as soil moisture, soil temperature, skin temperature, solar radiation, atmospheric pressure, and wind speed/direction at the standard 10-meter height. Eighty data values are transmitted via a GOES satellite transmitter in a 7-second burst within a scheduled 20-second time slot at the top of each hour. Data are available for processing at NOAA's National Climatic Data Center (NCDC) about 5 minutes past the hour.

NCDC conducts the primary functions of station installation, data ingest, processing, archival, and dissemination for the USCRN. NCDC employs an Oracle database to maintain both data and metadata. Station histories are documented using NCDC's Station History Archive database; image metadata, including photographs and scanned forms, are stored using NCDC's Web Search Store Retrieve Display. Operational problems are tracked using the USCRN Anomaly Tracking System from NOAA's National Environmental Satellite, Data, and Information Service (NESDIS). USCRN data processing software is written in FORTRAN and scheduling is provided through the use of crontabs.

A continuously running script ingests the incoming observations. If no data were ingested through the standard GOES data pathways, then a separate script will attempt to obtain the data through secondary pathways. As soon as data are received, they are replicated to a raw data archive on a mass storage device that also is backed-up off-site. Should data remain unsent from the remote site in any given four-month period, the site host is asked to download the observations (that are retained in memory for 5-6 months) onto a PDA. The PDA is returned to NOAA and data are downloaded to an FTP site for ingest.

Data ingest consists of checking the record length of the ASCII flat file and parsing the observations by time stamp. A FORTRAN common block is defined and the elements of each set of observations are stored in direct access files. (The record is discarded either if the record length differs from its anticipated length or if the position of numbers and decimal points of the elements are out of place.) After any duplicate observations are discarded, a script examines the hourly data stream to alert the network monitoring team (via email) about any missing data. Daily and monthly reports of missing data also are issued automatically.

Automated quality control (QC) is performed on the direct access files and then the data and resultant QC flags are loaded into an Oracle database. A script generates hourly, daily, weekly, and monthly QC reports that list the data failing any QC check. Members of the network monitoring team scan the QC reports and other products to analyze problems with the data or system. Flagged data are available online at http://www.ncdc.noaa.gov/crn/hourly.

The program that drives data ingest, QC, and output begins 5 minutes past the hour and takes approximately one minute to execute.

4.12.2 Analysis

The USCRN will provide the foundational data for climate research into the next century and currently is the model for all climate networks. The strengths of the USCRN include its redundant measurements at all remote stations, high standards of quality assurance, and execution speed for data ingest, QC, and dissemination. Because the USCRN is an end-to-end system, it can control the quality of the instruments in the field and, as a result, does not have experience with quality control of data from poorly maintained sites. In addition, the USCRN only measures temperature and precipitation; thus it has minimal experience with quality assurance of other environmental data. The USCRN is not focused on the continuous operations required by a real-time network. Although the USCRN will provide the data for the U.S. climate record during this century, there currently are no unique features of this network to warrant close attention by *Clarus*.

4.12.3 References

McGuirk, M., 2003: USCRN data management — ingest to access. Draft document, National Climatic Data Center, June 20, 2003.

4.13 WeatherShare

4.13.1 Description

System Owner: Western Transportation Institute

Reference Site: http://www.weathershare.org/

WeatherShare is a data integration and display system that was created by the Western Transportation Institute as a component of the Redding Incident Management Enhancement (RIME) program in California. The goal of the WeatherShare project is to streamline and integrate currently available road weather data in Northern California into a single source that can be accessed easily by incident responders and the traveling public.

Data currently incorporated for display on the WeatherShare website include observations from the California Data Exchange Center (CDEC), MesoWest, MADIS, and Caltrans. Agreements with these providers include the use of this information within a research project; therefore, commercialization of WeatherShare would require new agreements with these information suppliers. The website is publicly accessible (http://www.weathershare.org/) and also offers authorized users additional information and functionality. Data are mapped using HTML, JavaScript, and SVG (Adobe plug-in) technologies via the web site. Layers include observation networks (e.g., Caltrans data), observed variables (e.g., air temperature), and geographic overlays (e.g., streams, roads).

Queries by station and time interval output information in text and tabular formats. Users also may define threshold/alert values through a web interface, resulting in a map that only shows data values within the selected range(s).

Current and future work includes quality control development and evaluation, functionality for watches and warnings, user feedback/evaluation, and testing display capabilities on different client machines and web browsers. System expansion is expected to occur, incorporating more observation areas. Currently, no commercialized licensing exists, and given the nature of the existing agreements with the observation suppliers, this licensing status is not expected to change.

The WeatherShare system software uses Debian Linux (kernel 2.4.25), Apache version 1.3.33-2, MySQL version 3.23.49-8.8, Perl version 5.8.4-5, SSH version 3.4p1-1, and PHP4 version 1.3.10-2. System hardware includes a Dual Intel® Xeon™ 2.40GHz CPU, two 80-Gb hard drives, and 1 GB of memory.

4.13.2 Analysis:

The WeatherShare system provides important and timely data to DOT maintenance managers and the traveling public of northern California. It represents similar merged networks of RWIS and other weather stations. Unfortunately, as with most regional RWIS systems, the data processing system is proprietary and, as a result, it is difficult to assess system performance and scalability. There are no unique features of WeatherShare to warrant close attention by *Clarus*.

4.13.3 References:

Western Transportation Institute, 2005: WeatherShare Demonstration. Montana State University–Bozeman. Powerpoint Presentation, March 30, 2005, 27 pp.

4.14 West Texas Mesonet

4.14.1 Description

System Owner: Texas Tech University

Reference Site: http://www.mesonet.ttu.edu/

The West Texas Mesonet (WTM) is a joint program of the Atmospheric Science Group and Wind Science and Engineering Research Center at Texas Tech University. Initiated in 1999, the WTM currently operates 40 automated surface observing stations, two atmospheric profilers, and one upper-air sounding system across 28 counties. The primary objective of the WTM is to provide free, timely, and accurate meteorological and agricultural data to citizens of the South Plains/Rolling Plains region of western Texas (Schroeder et al. 2005). Measurements include air temperature, wind speed and direction, solar radiation, relative humidity, surface pressure, rainfall, soil temperature, and soil moisture.

Atmospheric observations are polled every 5 minutes and agricultural variables are measured every 15 minutes, resulting in 11,520 observations per day. Landline phone, cellular phone, extended line-of-sight radio, spread-spectrum radio, and the internet are used for two-way communications. Up to 78 days of 5-minute and 15-minute data are saved in the memory of the Campbell Scientific CR23X data logger. Data are available in text format, METAR, and Standard Hydrometeorological Exchange Format (SHEF) via a website, FTP, or LDM. Station metadata include station name, location, latitude, longitude, elevation, and several identifier codes.

Automated quality control is conducted through predefined tests applied to the collected data and examination of these test results by a “decision maker” to compute a final flag. In addition, the WTM staff provides qualitative checks on all incoming data to identify failing instruments, communication problems, or other issues affecting data quality and availability (Schroeder et al. 2005). A FORTRAN application was developed to apply the automated quality control tests based on those defined in Shafer et al. (2000). Parameters were set based on West Texas climatology, and additional or modified tests were used to conform to the geographical region. Each datum is flagged either “good,” “suspect,” “warning,” or “failure” based on the result of the range, step, persistence, like-instrument, and spatial tests. The data flags are maintained in a separate file from the raw data file.

Data are disseminated to users within a few minutes of the observation time for most stations. Images are generated for users by Unidata’s GEneral Meteorology PAcKage (GEMPAK) software.

4.14.2 Analysis

The West Texas Mesonet is an excellent example of a regional network operating with high standards on a limited budget. The strengths of the network include its use of several different automated QC tests, its minimal data latency from measurement to availability of quality assured data, and the varied data formats for use by the meteorological community. While adequate for its operations, the data processing system of the West Texas Mesonet was not designed as an object-oriented, modular system and likely does not scale well to the type of system that *Clarus* needs. Although the West Texas Mesonet is a quality environmental monitoring network, there currently are no unique features of this network to warrant close attention by *Clarus*.

4.14.3 References

Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, and J. W. Lipe, 2005: The West Texas Mesonet: A Technical Overview. *J. Atmos. Oceanic Technol.*, **22**, 211-222.

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **17**, 474-494.

5 COMMERCIAL ENVIRONMENTAL DATA COLLECTION SYSTEMS

It is very likely that many of the environmental data contributors to *Clarus* will be using COTS software to collect data from their ESS. As such, it is likely that *Clarus* will need to translate the environmental data products produced by those collector systems as part of its own data collection processes. This section discusses some of the prominent commercial environmental data collection systems.

5.1 LoggerNet™ Datalogger Support Software

System Provider: Campbell Scientific, Inc.

Reference Site: http://www.campbellsci.com/

LoggerNet™ is a data collection and network management package developed by Campbell Scientific, Inc. of Logan, Utah. The system is used specifically with dataloggers manufactured by Campbell Scientific. It can support a single station as well as networks of hundreds of stations. The software is intended to be an all-in-one application that allows for data collection and archiving, datalogger configuration and maintenance, network status monitoring, and graphical or tabular reporting of collected data.

LoggerNet utilizes client-server architecture. The main application is the LoggerNet server, responsible for all direct datalogger communication. The server also keeps track of any missing data and recollects data, if possible. Collected data are stored in a scalable data cache. Although not a requirement, the client-server architecture allows the LoggerNet server to run on a separate PC while client applications can run on remote workstations and connect to the server via TCP/IP. This also allows for distributed station load across multiple LoggerNet servers. The Oklahoma Mesonet, for example, utilizes four separate LoggerNet servers to manage over 150 stations.

The LoggerNet server supports connections to all Campbell Scientific datalogger types via direct serial, serial over TCP/IP, and Telephony Application Programming Interface (TAPI). Various communication systems can be implemented via these connections including VHF and spread spectrum radio, cellular telephone, and landline.

A suite of client applications is available for monitoring and configuring the server and for accessing collected data. Client applications do not directly access dataloggers, but instead request/transmit information through the server. Supplied clients include the Network Administrator for configuring network setup and data collection schedules; the Connect Screen client for individual datalogger maintenance such as program uploading and clock syncing; and the Status Monitor client providing graphical and textual reports of network status using information retrieved from the server. A LoggerNet SDK (Software Development Kit) is available for users who wish to develop specialized client applications and a command line interface for the server is available to facilitate network management via scripting. Several other client applications are available for exporting data to other applications/platforms/databases.

LoggerNet requires Windows NT, 2000, or XP with a minimum Pentium 2 processor with 128 Mbytes of RAM and 100 Mbytes of free disk space.

5.2 MetMan™ Network Software

System Provider: Vaisala, Inc.

Reference Site: http://www.vaisala.com/

MetMan™ Network Software is a data collection and network management platform from Vaisala, Inc. of Helsinki, Finland. In addition to supporting Vaisala-manufactured automated weather stations and observing systems, the system also may be configured to support some other makes of intelligent sensors, dataloggers, and automated observing stations. The MetMan system combines data collection services, a data management system, system configuration wizards, and a message management subsystem. The system is available in three configurations depending on network size, ranging from a single station to hundreds of stations. All services can reside on a single PC for the smaller network versions, while the large network version requires separate PC servers for communication services and data management.

The data collection services perform all station communications and support a large number of communications protocols including direct serial, satellite, TCP/IP, dial-up modems, GSM, SMS, and GPRS cellular, and various radio modems. Separate data collection schedules can be applied to blocks of stations or individual stations and both active (direct polling) and passive (station automatically sends data) collection is supported. The system also can track missing data, recollect data if necessary, and maintain station-clock synchronization across the network. The data services component also can collect data such as radar and satellite information from other computers.

The data management component provides permanent data storage allowing data to be stored in either text files or relational databases such as Oracle 9i. Several levels of data quality control are available to monitor for missing data and check measurement consistency so that erroneous data are not routed to users or other applications. These routines can perform internal, temporal, or spatial tests. The quality control routines also can be used to indicate required station maintenance. In addition to storing observational data, the database management systems can track and store station metadata.

Configuration wizards allow users to configure all system parameters such as data collection schedules and define/configure station sensor equipment. A message management subsystem is available for creating meteorological reports in many standardized formats while supplying automatic delivery to external users. MetMan requires Windows 2000 or Windows XP Professional.

5.3 SCAN

System Provider: Surface Systems, Inc.

Reference Site: http://www.ssiweather.com/

Surface Systems, Inc (SSI) of St. Louis, MO, has developed a data collection and management system known as SCAN. SCAN is designed for use with SSI's RWIS (Road/Runway Weather Information System) products. The platform is comprised of the SCAN Server, responsible for station communication, and a user interface, called SCAN Web, for viewing data.

The SCAN Server is the data collection engine for the SCAN platform and provides data polling at 15-minute intervals for the RWIS remote processing units. Several communication options are available, including CDPD, spread spectrum, LAN, WAN, direct wire, landline phone, radio, microwave, and fiber optics. Depending on the size of the server's hard drive, SCAN Server can receive data from 120 to 200 remote sites if they are on the LAN/WAN or about10 remote sites per dial-in line.

The SCAN Server ingests atmospheric and pavement sensor data, anti-icing system data, and video images from remote cameras and stores these data in a Microsoft Windows 2000 SQL Server database. The SCAN Server requires an Intel Pentium III 2.0 GHz processor, 512 Mbytes of RAM, and two 36 GB mirrored hard drives running Microsoft Windows 2000/2003 Server.

Data from RWIS sites are displayed as maps, graphs, or tables using SCAN Web. SCAN Web uses a standard web interface for displaying data collected from RWIS systems. SCAN Web assigns Uniform Resource Locators (URL's) to the RWIS data stored in the SCAN SQL Server database. Other products that can be viewed using the SCAN Web display tool include video from remote cameras (as still images or animations), station data histories, bridge sprayer status and operation/activation, and 24-hour pavement forecasts. SCAN Web requires an HTML browser such as Netscape Navigator 4.07 or later or Microsoft Internet Explorer 4.01 or later.

6 SYSTEM INFRASTRUCTURAL SOFTWARE

As with all contemporary software projects, the *Clarus* system implementation must rely on an underlying operating system to interface with the computer hardware. There are literally hundreds of different operating systems. Operating systems have been built to be general purpose, specific purpose, proprietary, and non-proprietary. Operating systems are just as diverse as the hardware on which they reside, having been created and becoming obsolete with each succeeding generation of computational resources; from mainframe, to minicomputer, to servers, to workstation.

The *Clarus* program should reside in a well-supported operating system. A custom operating system or specific purpose operating system would be expensive to develop, lack widespread support, and be unnecessary given the flexibility and robustness of general purpose COTS operating systems.

Microsoft Windows is the predominant operating system in the desktop computer market. UNIX, UNIX variants such as Apple’s OS X, and Linux are very common in the research and educational community. Linux and its various distributions are becoming increasingly popular with commercial industry.

Cost may or may not be an issue. Actual expense will be determined when the number of actual *Clarus* system deployments is decided. A distributed service topology will likely have more total hardware than a single installation.

There isn’t any function required by a *Clarus* system that cannot be supported by Microsoft Windows or a Unix/POSIX-compliant variant. If cost is an issue, an open source UNIX variant such as Linux would be the best solution. The meteorological industry is also very familiar with UNIX and Linux and will likely support a *Clarus* implementation built on a similar platform.

Another standard software package a *Clarus* system would use is a database. Like operating systems, there are many commercial and open source databases available. Oracle, Sybase, Microsoft SQL Server, Pervasive, and Microsoft Access are common commercial databases. PostGreSQL, Ingres, and MySQL are among the open source databases. All of these databases implement standard programming interfaces with varying support for stored procedures. Storage capacity and performance are parameters that need to be determined from a capability standpoint. As with operating systems, licensing cost may be an issue, depending on the number of deployed databases for a *Clarus* system deployment.

Beyond the operating system and database management system, there are various development platforms, toolkits, and frameworks (or, more generally, “middleware”) that purport to provide foundations for application development using standardized tools. Middleware has been useful in large institutional environments requiring sustained application development across the enterprise, but is generally inappropriate for individual applications. Most middleware packages are proprietary, and require specialized training for developers and application support personnel. Commercial packages are typically expensive to deploy.

For *Clarus*, there are no requirements that would drive a need for middleware. All of the assessment criteria in Section 3 can be met without incurring the acquisition and training costs of the middleware, and the proprietary nature of the solutions limits future development options. For all these reasons, middleware tools are inappropriate for *Clarus* development.

7 SYSTEM HARDWARE

The *Clarus* system hardware components are directly tied to both the system network architecture and the *Clarus* node software. Network topology affects both the hardware quantity and function and is, therefore, an integral part of this discussion. This section discusses some of the considerations and options that may affect hardware selection for *Clarus*, but leaves design decisions to the detailed design and deployment planning.

A distributed topology contains many *Clarus* installation nodes, each communicating with the contributors and service providers for which it is responsible. Storage capacity, processing power, and communication bandwidth are all dependent upon the data quantity collected and disseminated. More *Clarus* nodes generally equates to lower hardware requirements as there is more distributed hardware to spread out the information processing demand.

A single *Clarus* system installation responsible for collection, qualifying, and disseminating all environmental data and metadata for the entire North American Continent is a considerable undertaking. This will require enough communication bandwidth to handle 460 million incoming observations, at least 460 million qualified disseminated observations, and enough processing power, memory, and storage to qualify the incoming data every fifteen minutes or less.

A distributed *Clarus* system is similar to a centralized system in both interface and function. The difference between the two approaches is in the quantity of hardware at any given physical location. There are also some software differences to handle the various hardware cluster configurations. The desire for maximum redundancy, availability, and performance increases the complexity of an installation as well.

Maximum redundancy, availability, and performance can be achieved through hardware cluster implementations. Essentially, each primary function of the system is assigned to a group of two or more devices. This applies to power, communication connections, routing, switching, and load balancing, and also to data acquisition, processing, storage, and dissemination. This configuration compensates for up to half of all the devices to completely fail and still operate at a reduced capacity. The failed equipment can be replaced as soon as possible returning the system to its full capability.

In some situations, a completely redundant system is overkill in both initial expense and its ongoing maintenance cost. With a balanced system distribution, it is possible to meet a *Clarus* system's installation requirements with a single, two-server cluster in a fail-over configuration using hot-swappable RAID storage and their own uninterruptible power supplies, or installed in a facility with backup power availability.

The computer industry has experienced numerous changes over the years. The market has seen the demise of the mainframe and minicomputer that were replaced with parallel symmetric multiprocessor supercomputers, servers, and desktop workstations. Companies that once sold their own custom processors and interface hardware for their systems now sell systems built with standard parts. Market forces prompt processor manufacturers to continually put more capabilities into their products. One recent example is multiple processor cores packaged together. This constant pressure has made the line between workstation and server very fine.

The companies listed here are not intended as an exhaustive list, but are concrete examples of production computer components that adhere to industry standards. COTS computer hardware varies by:

- processor (Intel, Motorola, AMD, IBM);
- controller chipset (Intel, VIA, SiS);
- motherboard manufacturer (Asus, Tyan, IBM, Gigabyte, Leadtek, etc.);
- memory bus (DIMM, SDRAM, DDR, DDR2, Rambus) and bandwidth;
- interface bus architecture (ISA, EISA, PCI, AGP, PCIx); and
- storage interfaces (IDE, EIDE, SCSI, SATA).

Unless the *Clarus* program has a particularly special need, which is more likely with a single large installation, any individual *Clarus* system’s requirements should be able to be met with any readily available networking equipment and computer components and systems.

8 CONCLUSIONS

No existing system reviewed in this assessment meets the full intent of the *Clarus* specifications. None of the agency-sponsored systems are designed for use as a platform for independent systems development or for reuse of selected components. The commercially-sourced systems may fulfill some of the lower-level data collection tasks, but do not fit the quality control, dissemination, and administration needs of *Clarus*.

Several agency-sponsored systems provide *Clarus*-like capabilities and are in the process of being expanded in functionality or scope. The following systems merit further attention from the *Clarus* Initiative and may provide insight into specific implementation issues:

- MADIS,
- Mesowest,
- NERON,
- The Oklahoma Mesonet, and
- RWIN.

Commercially-sourced systems may be needed to collect data directly from environmental sensor stations and may be a cost-effective means of aggregating raw data from remote collectors. These systems do not, however, provide any native quality assessment capabilities, and are limited in their dissemination interfaces. If *Clarus* is designed to take information directly from the environmental data collectors, it will need to provide translators for each such collector.

COTS hardware and support software (such as operating system and database) will be adequate in all areas for a distributed *Clarus* implementation. A centralized *Clarus* system may require more specialized processing hardware to handle the expected load of data input, output, and quality control.

APPENDIX A - DEFINITIONS, ACRONYMS, AND ABBREVIATIONS

The following table provides definitions of terms, acronyms, and abbreviations to assist interpretation of this document. *The IEEE Standard Dictionary of Electrical and Electronics Terms* [B2], IEEE Std 610.12-1990, or IEEE/EIA Std 12207.0-1996 may be referenced for terms not defined here.

Term	**Definition**
AASHTO	American Association of State Highway and Transportation Officials
Acquirer	An organization that procures a system, software product, or software service from a supplier. (The acquirer could be a buyer, customer, owner, user, or purchaser.)
AMS	American Meteorological Society
API	Application programming interface. A well-defined set of functions commonly used by software to interface with libraries of reusable algorithms.
Architect	The person, team, or organization responsible for systems architecture.
Architecting	The activities of defining, documenting, maintaining, improving, and certifying proper implementation of an architecture.
Architectural Description (AD)	A collection of products to document an architecture.
Architecture	The fundamental organization of a system embodied in its components, their relationships to each other, and to the environment, and the principles guiding its design and evolution.
ASN	Abstract Syntax Notation
ASOS	Automated Surface Observing System
ATIS	Advanced Traveler Information Systems. A description for a group of interoperable services that enable dynamic management of transportation infrastructure and related activities. The project scope of the ATIS committee is to develop a minimum set of medium-independent messages and data elements needed by potential information service providers to deploy ATIS services, and provide the basis for future interoperability of ATIS devices.
CAP	Common Alerting Protocol. An open, non-proprietary standard data interchange XML format that can be used to collect all types of hazard warnings and reports locally, regionally and nationally, for input into a wide range of information-management and warning dissemination systems.
CCTV	Closed Circuit Television
Clarus	The *Clarus* system. An environmental data sharing system that collects, evaluates, and disseminates environmental data gathered from a geographically diverse set of environmental sensors.
CMML	Canadian Meteorological Markup Language
Collector	An electronic device used to convert environmental sensor electrical signals into environmental condition measured values and store them for retrieval.
ConOps	Concept of Operations

Term	**Definition**
Contributor	A managing agency or organization that owns and/or operates a set of environmental sensor collectors.
CSI	Cambridge Systematics, Inc.
DATEX	Data Exchange. A European standard effort for center-to-center data exchange.
DMS	Dynamic Message Sign
DOT	Department of Transportation
DSS	Decision Support System
ED	Environmental data. This data has not been processed by any quality checking algorithms.
EDR	An environmental data request sent to retrieve available environmental sensor information.
EM	Environmental metadata. Information about an environmental sensor station.
EMR	Environmental metadata request. A data request sent to retrieve information about environmental sensing stations.
ESS	Environmental Sensor Station
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
GRIB2	GRIdded Binary data format, Edition 2. An information format used to transmit grid-based weather forecasts from contributing offices to the NDFD and also one of the primary forms used to transmit the NDFD grids to weather information customers and partners
HAR	Highway Advisory Radio
HTML	Hypertext Markup Language
HTTP	Hyper Text Transfer Protocol. A communication standard for transmitting and receiving documents and other types of data over the Internet.
HTTPS	Secure Hyper Text Transfer Protocol.
ICC	(*Clarus*) Initiative Coordinating Committee
IMT	(*Clarus*) Initiative Management Team
in situ	From Latin, “in situ” means “in place.” As applied to meteorological data, it refers to data specific to a (fixed) point of observation.
ISP	Information Service Provider
ITE	Institute of Transportation Engineers
ITS	Intelligent Transportation System
ITS America	Intelligent Transportation Society of America
Life Cycle Model	A framework containing the processes, activities, and tasks involved in the development, operation, and maintenance of a software product, which spans the life of the system from the definition of its requirements to the termination of its use.
MADIS	Meteorological Assimilation Data Ingest System

Term	**Definition**
MDSS	Maintenance Decision Support System
MDT	Mobile Data Terminal
Metadata	In common information systems use, “metadata” is “data about data.” Within the meteorological community, this use has been extended to include data about objects related to weather observations. For example, location data for an ESS becomes metadata for the observation data.
MHI	Mixon/Hill, Inc.
MS/ETMCC	Message Set for External Traffic Management Center Communication.
NASA	National Aeronautics and Space Administration
NDFD	National Digital Forecast Database. A database supported by the National Weather Service that contains gridded forecasts of several ground-based weather elements such as temperature, humidity, and chance of precipitation.
NHI	National Highway Institute
NIST	National Institute of Standards and Technology
NOAA	National Oceanic and Atmospheric Administration. United States National Oceanic and Atmospheric Administration. A governmental administrative body responsible for managing programs and resources for weather and oceanographic science.
NTCIP	National Transportation Communications for ITS Protocol
NWOS	National Weather Observing System
NWS	National Weather Service
OCS	Oklahoma Climatological Service
OMB	Office of Management and Budget
PDA	Personal Digital Assistant
PMP	Project Management Plan
QC	Quality Checking.
QED	Qualified Environmental Data. Environmental data that has been evaluated by quality checking algorithms and contains a quality assessment flag.
QEDR	Qualified Environmental Data Request. A data request from environmental service providers and contributors to retrieve qualified environmental data from *Clarus* for value-added product delivery and quality feedback purposes.
RWIS	Road Weather Information System. Road Weather Information System. A unique system consisting of many meteorological stations strategically located alongside highways that allow the state Departments of Transportation to make more informed decisions during storms. Specialized equipment and computer programs monitor air and pavement temperature to make forecasts regarding how the weather impacts the operation and maintenance of the highways.
SAE	Society of Automotive Engineers. A group of engineers, business executives, educators, and students from more than 97 countries who share information and exchange ideas for advancing the engineering of mobility systems.

Term	Definition
SAE	Society of Automotive Engineers. A group of engineers, business executives, educators, and students from more than 97 countries who share information and exchange ideas for advancing the engineering of mobility systems.
SEP	System(s) Engineering Process
STWDSR	Surface Transportation Weather Decision Support Requirements
STWSP	Surface Transportation Weather Service Provider
System	A collection of components organized to accomplish a specific function or set of functions.
System Architectural Description (SysAD)	A collection of products to document a system's architecture.
TBD	To Be Determined
TCP/IP	Transmission Control Protocol/Internet Protocol.
TMC	Traffic Management Center
TMDD	Traffic Management Data Dictionary.
TOC	Traffic Operations Center
TRB	Transportation Research Board
UML	Unified Modeling Language
UTC	Universal Time Code
View	A representation of a whole system from the perspective of a related set of concerns.
Viewpoint	A specification of the conventions for constructing and using a view. A pattern or template from which to develop individual views by establishing the purposes and audience for a view and the techniques for its creation and analysis.
VII	Vehicle Infrastructure Integration
VSL	Variable Speed Limit
WIST	Weather Information for Surface Transportation
WRS	Weather Response System. A regionally-based service that electronically collects and processes weather forecast information into a coherent presentation for the purposes of traffic management and roadway maintenance.
XML	eXtensible Markup Language. A flexible text markup language used to create standard information formats that share both the format and the information to enable the interchange of structured data.