Project CRAFT: A Test Bed for Demonstrating the Real Time Acquisition and Archival of WSR-88D Base (Level II) Data

by

Kelvin K. Droegemeier and Jason J. Levit
School of Meteorology and Center for Analysis and Prediction of Storms
Norman, Oklahoma 73019

Kevin Kelleher
NOAA/National Severe Storms Laboratory
Norman, Oklahoma 73069

Tim D. Crum
NOAA/WSR-88D Operational Support Facility
Norman, Oklahoma 73069

Stephen A. Delgreco
NOAA/National Climatic Data Center
Asheville, North Carolina 28801

Linda Miller and David W. Fulker
University Corporation for Atmospheric Research
Boulder, Colorado 80307

Harry Edmon
University of Washington
Seattle, Washington 98195

DRAFT for submission to Bulletin of the American Meteorological Society

1 October 2000
Abstract

The US National Weather Service (NWS) recently completed the installation of 120 WSR-88D (NEXRAD) Doppler weather radars, concurrent with which was the deployment of 26 Department of Defense and 12 Federal Aviation Administration radars. Although these systems originally were designed without archival capability for the full-volume, full-precision Level II (base) data, an interim solution, based upon 8 mm tape technology, eventually was implemented. Since 1992, the National Climatic Data Center has been archiving the data tapes and using the same media to provide base data to the national community. Although an innovative and satisfactory solution several years ago, the tape system is extremely human-resource intensive, costly, inefficient, and unreliable.

In an attempt to begin addressing the long-term needs for WSR-88D base data archival, and in light of the compelling need for real time access to base data for use in a variety of applications, especially numerical weather prediction, the Center for Analysis and Prediction of Storms at the University of Oklahoma joined forces in 1998 with the University Corporation for Atmospheric Research, the University of Washington, the National Severe Storms Laboratory, and the WSR-88D Operational Support Facility to establish the Collaborative Radar Acquisition Field Test (CRAFT). The principal goal of CRAFT is to demonstrate the real time compression and internet-based transmission of WSR-88D base data from multiple radars with a view toward nationwide implementation. The initial test bed of six radars, located in and around Oklahoma, has been delivering real time base data since early 1999 with virtually no interruptions, and now is being expanded to include some 30 additional radars.

We describe in this paper the motivation behind Project CRAFT as well as the technical implementation strategy and potential impacts on research, operations, and education. Additionally, we discuss opportunities for including other surveillance radar systems, especially
those operated by the Federal Aviation Administration, as well as the role of Abilene and Internet2 in taking the CRAFT concept to the entire NEXRAD network.
1. Introduction

The US National Weather Service (NWS) recently completed the installation of 120 WSR-88D (NEXRAD) Doppler weather radars (Alberty et al. 1991) in a national network (Figure 1) as part of its $4.5 B Modernization and Associated Restructuring Development Plan. (Another 26 Department of Defense and 12 Federal Aviation Administration NEXRAD Radars also were installed, for a total of 141 systems in the continental United States.) This unique observing system provides nearly continuous single-Doppler radar coverage across the continental United States which, when coupled with the radar's superb sensitivity, sophisticated processing algorithms, and advanced user training, has led to substantial improvements in the identification and short-term warning of hazardous local weather (e.g., Crum et al. 1998); to new opportunities for quantitative precipitation estimation and improved flood forecasting (e.g., Foufoula-Georgiou and Krajewski 1995; Fritsch et al. 1998; Droegemeier et al. 2000); to a deeper understanding of mesoscale weather and opportunities for its explicit numerical prediction down to the scale of individual thunderstorms (e.g., Lilly 1991; Droegemeier 1997), and to interesting non-meteorological applications of radar data (e.g., Gauthreaux and Belser 1998).

Although the WSR-88D radars originally were designed without archival capability, the value of their unique observations for scientific research soon was recognized. An interim solution for the recording and long-term archival of full-volume, full-precision Level II data (also known as base data; see section 2) data was achieved by outfitting each radar with an 8 mm robotic tape cartridge recording system (Crum et al. 1993). Since 1992, the National Climatic Data Center (NCDC) has been archiving the 8 mm tapes and using the same media to provide base data to the national community. As described in section 2, the archival and re-distribution process -- though an innovative and satisfactory solution several years ago -- is
extremely human-resource intensive, costly, inefficient, and unreliable. Rapid advances in data transmission, compression, and storage technologies suggest that far better solutions are now available and can be implemented with high reliability at reasonably low cost.

In an attempt to begin addressing the long-term needs for WSR-88D base data archival, and in light of the compelling need for real time access to base data for use in a variety of applications, especially numerical weather prediction (e.g., Droegemeier 1997; Dudhia et al. 1999), the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma joined forces in 1998 with the University Corporation for Atmospheric Research (UCAR), the University of Washington, the National Severe Storms Laboratory (NSSL), and the WSR-88D Operational Support Facility (OSF) to establish the Collaborative Radar Acquisition Field Test (CRAFT; Droegemeier et al., 1999). Funded initially by a grant from the Oklahoma State Regents for Higher Education and now by the National Oceanic and Atmospheric Administration (NOAA), CRAFT is an experiment in the real time compression and internet-based transmission of WSR-88D base data from multiple radars. The initial test bed of six radars, located in and around Oklahoma, has been delivering real time base data since early 1999 with virtually no interruptions, and now is being expanded to include some 30 additional radars to serve as the prototype for a national system.

This paper provides an overview of Project CRAFT and its potential impacts on research, operations, and education. Section 2 details the types of and uses for WSR-88D data, as well as limitations associated with the current NCDC base data archival methodology. The technical aspects of Project CRAFT are detailed in section 3, along with our experiences to date in collecting base data via the commodity internet. We discuss in section 4 efforts now underway to expand Project CRAFT via the high-speed Abilene network and Internet2 project,
and in section 5 outline plans for compatibility with the NEXRAD Open Systems Architecture.

A summary and view toward the future are offered in section 6.
2. WSR -88D Radar Data: Types, Availability, and Utilization

a. Data Types and Long-Term Archival Procedures

Raw data collected by the WSR-88D radar are used to generate four so-called Archive Level data sets (OFCM 1991).

The first, or Archive Level I, consists of the analog, time domain output from the radar receiver (also known as meta data) and includes information concerning radar synchronization, calibration, antenna position, and operational mode. Archive Level I data typically are used for signal processor optimization studies.

Archive Level II (also known as base) data are full precision, full spatial resolution data of the three Doppler moments (radial wind, reflectivity, and spectrum width), along with appropriate meta data. (Hereafter we use the terms Level II data and base data interchangeably.) As described more fully below, these data are valuable for scientific research and also are utilized for training, operations support, and algorithm development.

Archive Level III data (also known as NEXRAD Information Dissemination Service, or NIDS) are a series of 24 products generated by the radar product generator (Alberty et al. 1991). Under a term agreement with the National Weather Service, several private vendors have for several years been distributing NIDS data to a broad user community. The distribution agreement will expire at the end of 2000, and thereafter all NIDS data will be available free of charge, in real time, from a centralized NWS server.

Finally, Archive Level IV data consist of any information or products that are available at the user display station (known as the Principal User Processor, or PUP, it now has been replaced by the Advanced Weather Information Processing System, or AWIPS; e.g., McCarthy 1999).
The NCDC in Asheville, North Carolina is responsible for collecting and archiving all Level II and Level III data from both civilian and military WSR-88D systems. The process for Level III data, shown schematically in Figure 2, starts with the creation of WORM (write once, read many times) compact disks at the radar site (note that Level III products are archived only for the NWS radars). Each disk, which contains up to 600 megabytes of data, is checked upon receipt by the NCDC for errors and missing data. Once cataloged, the contents are transferred to the hierarchical data storage system (HDSS) and the original disk archived.

On average, the NCDC receives about 4 gigabytes of Level III data each day, or about 1.5 terabytes annually. By late 1999, the NCDC had archived some 13,000 optical disks, with inventories on the HDSS dating to 1994. The archive is 89.5% complete for all 120 NWS radars, and as shown in Figure 3, orders for Level III data come predominantly from consulting meteorologists for use in forensic studies such as aircraft and boating accidents.

In order to cover personnel and related costs associated with filling radar data requests, the NCDC charges $100 for 24 hours worth of Level III data for a single station. Data may be received via computer tape or ftp. Color hard copies cost $15 per image. Some customers, such as those affiliated with the three NEXRAD agencies, receive the data free.

The archival of Level II data is substantially more complicated owing to the greater volume of information (the NIDS base reflectivity and velocity products utilize only the lowest four elevation angles out of a possible 11 or 14 tilts, based upon the mode of radar operation) and the different recording hardware. Figure 4 shows a schematic of the Level II archival process, which begins when a radar site mails a set of ten 8 mm tapes to the NCDC. Each tape, which may contain up to 4.8 gigabytes of data, is checked for errors, given a header, and reblocked to 31,616 bytes. The data are then merged and compressed, at a rate up to 10:1, to
another archive tape (IBM 3590) for long-term storage (note that the data are not written to the HDSS). The original 8 mm tapes received from the radar site are then recycled.

On average, the NCDC receives 165 gigabytes of Level II data each day, or about 60 terabytes annually. By late 1999, the NCDC had archived some 42,000 8 mm tapes, with some data available as early as 1989 but most files in the archive dating to 1995. As of late 1999, the Level II archive for all 120 NWS radars was 64.9% complete, in comparison to 35.7% and 9.5% for the Department of Defense and FAA radars, respectively. Consequently, for the entire 158 radar network, the archive is only 55.7% complete (we examine the reasons for this in the next sub-section). As shown in Figure 5, orders for Level II data come predominantly from NOAA and the NWS, or from groups working with them. University research accounts for slightly less than 12% of the requests, and as described further below, this number would likely be significantly higher if the data were more readily available.

As is the case for Level III data, the NCDC must cover its personnel and related costs in distributing archived Level II data to the user community. For customers not covered by the NEXRAD free data policy (i.e., non-NEXRAD agencies), the first 48 hours of data from a single radar site costs $100, and each subsequent 24-hour period from the same site costs $25 (an $11 service and handling charge also is assessed). Data are provided only on 8 mm compute tape. From July, 1993 through November, 1999, 9,020 of the 9,694 Level II requests filled -- or 93% -- were done so free of charge.

b. Limitations in the Archival and Acquisition of Level II Data

As noted in a recent National Research Council report (NRC 1999), the NEXRAD radar system is a unique observing tool that provides data of unprecedented spatial and temporal resolution. The NRC article also points to the need for making NEXRAD data more readily available to the operational, research, and educational communities, including real time access,
and for modernizing and streamlining the processes by which historical data sets are archived at
and retrieved from the NCDC. Before examining how Project CRAFT is attempting to help
address these issues, a discussion of the limitations of the current archival process is warranted.

As noted in the previous sub-section, Level II data are collected at each radar site on 8
mm tape. Neither the recording media nor the hardware were designed for the stresses to which
they are subjected in the field. Consequently, the tape system failure rate is high and partially
explains the low archive rate noted above.

The slow speed of the 8 mm tape drives, 92 of which are presently maintained at the
NCDC, lengthens significantly the time needed to both read and write Level II data.
Consequently, an order for even a modestly large data set involving a dozen radars may take a
few months to be processed, and the cost for the data, especially for use in long-term, large-area
climatological studies, can run into the hundreds of thousands of dollars -- far more than can be
accommodated by a typical research grant. Furthermore, no mechanism presently exists to
peruse Level II data online or mine vast amounts of historical data, and the cost to maintain the
now outdated tape system, both in the field and at the NCDC, exceeds $0.75M per year.

As a result of these and other limitations, the non-NOAA scientific community,
especially at universities, has largely avoided using archived Level II data for either research or
education (Figure 5). In other words, the low usage of Level II data by the academic
community in Figure 5 does not reflect a lack of interest, but rather a lack of cost-effective
access.

c. Examples of Level II Data in Research

To underscore the value of WSR-88D base data in research, we present a few illustrative
examples from recent studies.

(1) Storm-Scale Numerical Weather Prediction
Exploration of a new use for WSR-88D base data began in 1989 when the National Science Foundation created CAPS at the University of Oklahoma. The mission of CAPS is to demonstrate the practicability of using numerical models, operated at resolutions down to approximately 1 km, to predict explicitly the initiation and evolution of local, high-impact weather using the WSR-88D radar as a principal data source.

Because a single Doppler radar provides only the radial component of the wind (with all three components being required for initializing a forecast model), and because the coverage patterns of the WSR-88D radars overlap only over small areas at rather large ranges, CAPS developed several techniques to retrieve the unobserved polar and azimuthal wind components from time series of single Doppler data (e.g., Sun et al. 1991; Xu and Qui 1994a,b,c; Qui and Xu 1992, 1996; Shapiro et al. 1995; Zhang and Gal-Chen, 1996). These methods, and extensions of them developed by others (e.g., Sun and Crook 1994, 1997), have been used with notable success in case studies (e.g., Lazarus 1996; Weygandt 1998; Crook 1994) as well as real time testing (e.g., Janish et al. 1995; Droegemeier et al. 1996a,b; Xue et al. 1996a,b; Carpenter et al. 1999; Sun and Crook 1999).

Since 1994, CAPS has been evaluating its storm-scale numerical weather prediction system in real time (see http://caps.ou.edu/wx) in collaboration with the National Weather Service, the Storm Prediction Center, and other entities of the Federal and private sector. Unfortunately, the model initialization procedure generally has been limited to the use of NIDS data which, although available in real time for every WSR-88D radar, are generally of limited value for numerical prediction by virtue of their reduced volume coverage and substantially degraded precision.

To illustrate the importance of base data in storm-scale numerical weather prediction, we show results from two high-resolution forecasts in which the CAPS Advanced Regional
Prediction System (ARPS; Xue et al. 1995; Xue et al. 2000a,b) was initialized first with NIDS data and then with base data. Figure 6 shows a mosaic of radar reflectivity at 6:40 am CDT (1240 UTC) and 10:43 am CDT (1643 UTC) on 5 April 1999. The convective line bisecting the Red River formed along a low-level north-south boundary and moved rapidly eastward, eventually extending southward into southern Texas. Although this event was relatively strongly-forced and thus, in principle, should be less dependent upon small-scale observations, the results indicate just the opposite.

Using the National Centers for Environmental Prediction (NCEP) 1200 UTC Eta forecast as the background state, wind profiler, rawinsonde, and surface observations, the latter from both the standard network as well as from the Oklahoma Mesonet, were assimilated into the ARPS (see Brewster 1996 for details of the ARPS assimilation system). Level III (NIDS) reflectivity data from several nearby radars, including those at Oklahoma City and Fort Worth, also were added via a diabatic initialization and vertical velocity adjustment scheme in order to provide information on the convective scale (Zhang and Carr, 1996; Zhang 1999). Although methods have been explored for using NIDS radial velocity data in model initialization (Levit and Droegemeier 1999), they were not used here.

The 700 hPa vertical velocity field for the final assimilated state containing NIDS data, shown in Figure 7a, depicts only a broad region of rising motion (red) ahead of the convection and subsidence (blue) behind. The maximum updraft is 0.09 m s\(^{-1}\), and completely absent is any velocity structure of the convection itself. Although the moisture field from the surface to around 700 hPa (not shown) was enhanced in the region of active convection, the associated 3-hour forecast, the boundary conditions of which were provided by the Eta model, virtually misses the deep convection in southeastern Oklahoma (Figure 7b), and also underestimates the
intensity of convection in eastern Texas. (Note that the 3 km spatial resolution used precludes capture of the fine details shown in the radar data.)

A forecast identical to that presented above was performed with Level II reflectivity and velocity data from the Oklahoma City and Fort Worth WSR-88D radars replacing the Level III (NIDS) data. The Shapiro et al. (1995) single Doppler velocity retrieval (SDVR) procedure was applied to retrieve from the radial wind data the azimuthal and polar components, followed by a Gal-Chen (1978) thermodynamic retrieval to obtain the temperature and pressure fields and the Zhang and Carr (1996) moisture adjustment. Details of the initialization procedure are found in Weygandt (1998).

As show in Figure 8a, the analyzed 700 hPa vertical velocity field obtained using base data contains a clear signal of the convective-scale updrafts and downdrafts. The maximum updraft is $0.33 \text{ m s}^{-1}$, which by convective storm standards may seem small, but is nonetheless significant given the 3 km horizontal resolution used as well as the spatial extent of the updraft itself. The 3-hour forecast based upon this initial condition, shown in Figure 8b, better captures the intense convection in southeastern Oklahoma (compare with Figure 7b), though the coarse grid resolution precludes a faithful representation of the details.

Although NIDS data have proven useful for storm-scale model initialization in some situations, they suffer from two limitations that preclude general use, especially in weakly-forced situations. First, only the lowest four tilts of the volume scan are transmitted, which limits the data content depending upon the location of the radar relative to the feature of the interest. Second, and more importantly, the data are quantized in discrete bins, e.g., a radial wind of 13.5 kts would be assigned a value of 10 kts because it falls into the [10,20) bin. This degradation of precision introduces errors and imbalances that have a deleterious impact on forecast quality.
(2) *Storm-Scale Variability*

The ability to assimilate observations into numerical prediction models via variational approaches depends upon knowing the natural variability of each observed field (e.g., Schlatter et al. 1999). Although such variability has long been established for the large scales of the atmosphere, it currently remains elusive for individual convective storms as well as their larger aggregates and wintertime counterparts. This limitation must be overcome by the time storm-resolving models (e.g., Dudhia, 1998) are implemented operationally by the NWS.

WSR-88D base data, if made readily available from each site, could be processed, perhaps in real time, to help establish the natural variability of the atmospheric mesoscales. Numerous studies have long sought to address this issue (e.g., Lilly 1983a,b), in many cases using high quality data but with rather severe limitations in spatial extent, representativeness, and temporal record.

(3) *Synthetic Storm Feature Climatologies*

Concurrent with the emplacement of the WSR-88D radar network was the initiation of a significant effort to develop algorithms for the automated detection of severe weather hazards (e.g., Johnson et al. 1998; Witt et al. 1998; Stumpf et al. 1998; Mitchell et al. 1998). Many algorithms based upon NIDS data are now operational in AWIPS, and stand-alone systems that utilize base data, such as the NSSL Warning Decision Support System (WDSS; Eilts et al. 1996), have yielded significant improvements in forecast skill (see references cited above).

The availability of these and other algorithms which, owing to the need for rapid forecaster response during severe weather are computationally very efficient, makes possible the creation of "synthetic storm feature climatologies" if applied to streams of WSR-88D data, especially in real time. The ability to continuously catalog, on a national basis, the occurrence (based upon the algorithms) of hail, mesocyclones, tornadoes, vertically integrated liquid, storm
motion, and other storm attributes would be of inestimable value to meteorologists, hydrologists, urban planners, power and communication utilities, the insurance industry, the military, and numerous other groups. Indeed, such studies based on limited NIDS data (e.g., Pugh 1999) already have shown their value, and considerable benefit would accrue from the processing of base data on a national scale for quantitative precipitation estimation in hydrologic applications (e.g., Fritsch et al. 1998; Droegemeier et al. 2000).

d. Use of Level II Data in the Classroom

At the present time, WSR-88D data are available to the academic community only in the form of Level III or NIDS products. Although very timely, these data limit the ability of the student to create special products or manipulate the data in ways that would enhance understanding, especially on topics related to radar geometry and data processing. Level II data, on the other hand, offer boundless opportunity, especially when coupled with real time decision support systems such as the WDSS.
3. An Internet-Based Prototype for Collecting and Archiving Base Data

   a. **System Configuration**

   The foundation for Project CRAFT was laid several years ago when, as part of its efforts to develop radar-based warning decision support tools using Level II data, the NSSL created the Radar Interface and Data Distribution System (RIDDS; Jain and Rhue 1995, http://www.nssl.noaa.gov/srad/mss/mssproducts.shtml) to extract base data from the NEXRAD radar product generator (RPG) in real time.

   As shown in Figure 9, the RIDDS consists of Sun SparcStation 5/110 workstation, running special software developed by the NSSL, that takes base data from the Wideband 3 port and outputs it as UDP packets to an 8 port ethernet hub. A circular buffer on the RIDDS workstation holds about one volume scan of data. Until the advent of Project CRAFT, a maximum of 8 users could access the RIDDS base data by establishing a (usually rather expensive) T1 network connection (1.55 megabits/sec) to the ethernet hub and connecting to it an appropriate computer (Figure 9). By the end of 1999, some 40 RIDDS sites were active around the nation (Figure 10), mostly for testing the WDSS.

   Project CRAFT sought to enhance the above capability by a) using the internet as a framework to provide real time base data access to a virtually unlimited number of users; b) reducing significantly the speed, and thus the cost, of the network connection to the RIDDS computer by implementing a real time loss-less compression algorithm (i.e., one that outputs exactly the same data after de-compression, without any loss of information); and c) evolving a system that could be scaled to accommodate an arbitrary number of radars as well as larger data streams owing to future radar enhancements (e.g., dual-polarization, phased-array). The basis for doing so was a test by the University of Washington in which an inexpensive personal computer, running the UCAR Unidata Local Data Manager (LDM) software, was linked to the
RIDDSS system and used to compress base data in real time and then transmit it to local users via the internet.

The LDM software (Rew and Davis 1990; Davis and Rew 1994; UCAR 1998), which is used by numerous universities as well as federal government laboratories and operational centers, is a distributed system for event-driven data distribution that acquires data from and shares them with other networked computers (Figure 11). LDM can handle data from the NWS "NOAAport channel 3" data streams, including gridded data from numerical forecast models, as well as WSR-88D NIDS products, lightning data from the National Lightning Detection Network, and GOES satellite imagery.

Data passed to an LDM server are processed in a variety of ways depending upon data identifiers and a configuration file. The LDM server can be instructed to append a particular product to an existing file, execute an arbitrary program with the data product as input, store or retrieve products in a simple key-access database, or forward the data along to other running servers. This latter capability is especially important for Project CRAFT as it provides redundancy in the network.

The CRAFT system utilizes the University of Washington concept by linking to the RIDDSS workstation an inexpensive (approximately $1,500) personal computer (PC) running the Linux operating system and LDM (Figure 12). This PC requires only nominal memory (64 megabytes) and disk space (2 gigabytes), and is purchased without a monitor. The data then travel from the PC to an optional $2000 multi-port router and onto a standard 56 kilobit/sec (56K) phone line (the router allows for installation of redundant 56K lines).

A key aspect of CRAFT is the addition to LDM of an off-the-shelf, public domain, lossless data compression algorithm known as BZIP2. It compresses the base data, in packets of 100 radials, down to an average of 1/12th their original size. (The maximum compression
achieved is about 16:1 for clear air data, and the minimum to date in heavy weather has been about 8:1.) In light of the fact that a single 5 or 6-min volume scan is typically a few megabytes in size and never exceeds about 15 megabytes (i.e., a volume scan completely filled with echoes), the data compression achieved is more than adequate for a 56 K phone line. Indeed, the aggregate compressed base data rate for the entire national WSR-88D network is only 30-40 megabits/sec, and consequently bandwidth is not the principal issue with regard to acquisition (see section 4). At the end user site (Figure 12), another computer running LDM ingests and decompresses the base data, again in real time. In the event of a communications failure, the LDM PC at the radar site automatically stores and can retransmit up to 4 days worth of base data. This amount can be increased simply by increasing the capacity of the local PC disk.

Two other significant features of LDM add to the flexibility of Project CRAFT. The first, scalability, is clearly evident in Figure 11, which shows that an effectively unlimited number of both "data source" as well as recipient nodes can be added because each functions as a transmission point. In addition, a single LDM node can transmit to multiple other LDM nodes. As shown below and discussed further in section 4, this capability allows data from the present CRAFT radars to be transmitted simultaneously to the University of Oklahoma, the NSSL, and the NCDC.

b. System Performance and Reliability

By early 1999, six WSR-88D radars in and around Oklahoma (Figure 13) were delivering base data in real time using the methodology described in the previous sub-section. The installation of the LDM PC at each associated NWS office (the location of the RPG) took only a few hours, with 56K phone line charges ranging from $2000 to $6000 per year depending
upon local tariffs and area code boundaries. Special rates were negotiated where possible in collaboration with local providers.

Assessing the reliability of base data transmission is one of the key goals of Project CRAFT, and Figure 14 shows a simple real time monitoring system that tracks network health by plotting the real time data rate in bytes per second. The upper time series spans a window of roughly a day, while the lower one covers an entire week. As noted toward the top of the image, this particular 56K line to the Amarillo, Texas WSR-88D radar had been up for 154 days without interruption. Similar displays have been created by NASA for the Hytop, Alabama radar (Figure 15), which uses the CRAFT acquisition strategy. As discussed further in section 4, data of the type shown in Figures 14 and 15 are valuable for studying the networking characteristics of radar data transmission, especially "burst rates" that help define quality of service requirements.

Although personnel constraints have precluded the creation of detailed statistics regarding the percentage of data received from the six CRAFT radars, we believe the figure to be well over 95 percent. As described in the next section, the NCDC is now ingesting data from a dozen CRAFT radars and not only has verified the data to be identical to those on the 8 mm tapes, but also is collecting reliability statistics.

c. Archival Tests at the NCDC

One of the principal goals of Project CRAFT is to demonstrate the capability of directly ingesting and archiving base data at the NCDC, thereby eliminating the expensive and unreliable 8 mm recording technology now in use (see section 2a).

In early June, 2000, the NCDC began receiving compressed base data in real time from the original six CRAFT radars (Figure 13) via the commodity internet. By early August, these data were being transferred directly, and automatically, to the NCDC mass storage system
(Figure 16). In September, 6 other radars were added (see section 4). The base data from CRAFT have been verified to be identical to those from the 8 mm tapes, though in contrast to the tapes and owing to the present design of the RIDDS, the real time data stream does not transmit meta data (see section 1). This limitation will be overcome with the move to the open systems radar architecture, as described in section 5.

Even though the present ingest system at the NCDC utilizes a standard T1 line, the data from the CRAFT radars utilizes only a small fraction of available bandwidth. Efforts are now underway to develop and test an automated retrieval system whereby users can query the NCDC data base and download compressed, archived data via the internet. Additionally, visualization software is planned to be made available so that users can browse files on the web prior to download.

d. Benefits and Limitations

One of the principal benefits of the CRAFT methodology, particularly in terms of cost and reliability, is its utilization of existing software (e.g., LDM, BZIP2, Linux) and networking technology (the internet). As described further in section 4, this hardware-software framework lays the foundation for the cost-effective expansion of CRAFT to all WSR-88D systems; provides for the inclusion of other radars (e.g., from the FAA); accommodates the transmission of base data in real time to virtually any user in the world; and allows expansion to larger data sets via enhancements to the existing radars (e.g., dual polarization capability and more dense scanning strategies) or the introduction of new ones (e.g., phased-array).

With regard to long-term base data archival at the NCDC, the CRAFT system has reduced the number of processing steps from 8 to 2 and the time required to archive a data set by a factor of 4. In addition, by eliminating the expensive and unreliable 8 mm tape recorders, the CRAFT concept eventually will eliminate the time spent by field personnel working with
them, eliminate the associated hardware maintenance costs, and also substantially reduce the requirements for tape media. The data archival rate, which now stands at 65% for the NWS radars, should approach 100% with the CRAFT system, and the direct transfer of base data to the NCDC HDSS, along with the planned development of software for user data retrieval via the web, will reduce dramatically the time required to obtain historical data.

The CRAFT system is, of course, not perfect. Initial problems with the 56K phone lines, due mostly to the use of multiple telecommunication providers, were quickly overcome. As noted above, the lines are quite stable, though they do represent a single point of failure (see next sub-section). The RIDDS computers and Linux PCs also have been quite reliable, and both are managed via remote log in (only one trip during an 18 month period was needed to replace a malfunctioning PC, and the installation was accomplished in less than an hour). However, the present CRAFT system has no provision for redundancy in local communications or computer hardware, and these issues are being addressed in the next phase of the project (see section 4).

Finally, although the internet itself might be viewed as unreliable for use in time-sensitive meteorological operations, especially because users or agencies lack ownership of or control over every link involved, our experience to date suggests just the opposite. Indeed, although major internet disruptions are indeed possible (e.g., inadvertent excavation of primary trunk lines), with security an important issue as well (see section 4), the internet is so pervasive and so redundant that it probably represents one of the most reliable means of data transmission currently available.
4. CRAFT-2: Toward a National Base Data Acquisition System Based on Internet2 and the Abilene Network

a. Motivation

The original six radars in Project CRAFT (Figure 13) were intended to demonstrate the basic capability of compressing and transmitting WSR-88D base data in real time via the internet. Expansion of this concept to the entire radar network will require not only additional bandwidth, but more importantly a variety of security and quality of service capabilities not presently available from the commodity internet (e.g., reserved bandwidth and bandwidth on demand). These and other features are essential if prediction and warning systems that utilize base data from CRAFT are to meet operational reliability standards.

To illustrate the need for a dynamic, intelligent network that responds to loads that change with time, one need only look to the future of numerical weather prediction. Figure 17a shows a schematic of the type of numerical forecast system now being prototyped by CAPS (Droegemeier 1997) and developed for the NWS (Dudhia et al. 1998). By using a series of nested grids, each having successively higher resolution, the model is, in principle, capable of capturing the details of local, high-impact weather while at the same time representing the larger-scale environment. With automated nesting, in which new grids are initiated over regions of developing weather and removed over regions where active weather is dissipating, the computation becomes even more dynamic because the model configuration changes with time (Figure 17b).

By the time such capabilities become operational, it will be practicable to run part of the calculation (e.g., the outer grids) at one computing center and another part (e.g., the inner grids) at a center perhaps thousands of miles away in a manner that maximizes throughput and load balances the entire computation (a concept known as the "virtual machine room"). And of
course, the forecast model must have real time access to base data from the entire national network of WSR-88D radars, the networking quality of service for which must also change in response to the weather (i.e., a landfalling hurricane would require high quality of service and bandwidth in a particular region of the country for a particular period of time). Indeed, this concept of dynamic networking linked to dynamic computation and data acquisition in numerical weather prediction is the embodiment of the national computational grid (Foster and Kesselman, 1999).

b. System Design

Exploration of the quality of service issues described above, and the extension of CRAFT to the entire WSR-88D network, are being explored within the context of two national networking infrastructures: Internet2 and Abilene. Internet2 is a non-profit consortium of over 170 universities in the United States (Figure 18) that develops and deploys advanced network applications and technology. Abilene (Figure 19) is a gigabit/sec backbone that supports Internet2 activities by providing a high-speed interconnect among all collaborators. Both Abilene and Internet2 are managed by the non-profit University Corporation for Advanced Internet Development (UCAID).

An overlay of Internet2 universities and the WSR-88D network (Figure 20a) reveals that the average distance from any radar to the nearest Internet2 network node is only 56 miles. By running a 56K phone line from each radar to the nearest Internet2 node (Figure 20b), in some cases via reduced-rate agreements with regional networks, as was the case in Oklahoma, the CRAFT concept is immediately reproducible, at low cost, nationwide. Once the base data arrive at an Internet2 site, they immediately reside on the high-speed Abilene backbone and thus are available to all users, even those on the commodity internet. To access data from a particular radar, a user having an LDM system would simply enter the appropriate radar address.
The logical networking topology envisioned for this concept centers around the transmission of base data from all radars across the Abilene backbone (Figure 21). Selected sites on the Abilene ring, such as high-priority NOAA facilities, NCAR, and some universities, would serve as transfer points for the entire data stream, while "satellite" nodes not linked directly to Abilene (e.g., universities, private companies, other government facilities) would obtain base data via the commodity internet, or Abilene if they are so connected, for as many radars as available bandwidth allows. As networking capabilities continue to expand, bandwidth limitations on the commodity internet are likely to vanish entirely.

c. Initial Tests and CRAFT-2

The concept described above is being tested as part of the next phase of Project CRAFT, known as CRAFT-2. In particular, approximately 30 RIDDS-equipped radars are being added to the present CRAFT network (Figure 22), in many cases with line charges from the radar to the Internet2 point of presence paid by ongoing research projects. This broad geographic distribution will provide an acid test of overall reliability, network efficiency, and real time archival at the NCDC and eventually ingest at the NCEP.

At the present time, four additional radars associated with the Salt River Project in Arizona (Dempsey et al. 1998) have been added, along with radars at Seattle, Washington and Rapid City, South Dakota. The NCDC is now archiving data from all 12 CRAFT-2 systems via the commodity internet, and is in the process of establishing a connection to Abilene. As part of this effort, NSSL and the University of Oklahoma are working to improve the data compression algorithms, and exploratory work also is underway on interactive data mining.
5. The Move to Open RPG

Because Project CRAFT is a proof-of-concept activity, it makes use of technology, like RIDDS, that fulfills present needs but does not have a long-term future in the WSR-88D program. If the CRAFT concept is to be implemented in a stable manner for future operations, it must be fully integrated into the WSR-88D framework. The opportunity to do so now exists by virtue of a long-standing effort within the NEXRAD Product Improvement Program to improve overall system functionality, reduce cost, and provide greater flexibility for the future. Known as the Open Systems Architecture (Saffle and Johnson 1999), this project involves replacing existing hardware and software with off-the-shelf "open" components that make use of industry standards such as unix.

In the new Open Systems Architecture, the deployment of which is scheduled from June, 2001 through December, 2002, the functionality of the RIDDS and LDM computers will be replaced by a single machine known as the base data distribution system (BDDS). Consequently, efforts in CRAFT-2 now are being directed toward examining the integration of LDM and improved data compression algorithms into the BDDS, with testing anticipated by spring 2001. Although the time schedule for BDDS deployment initially will preclude a CRAFT-type of data ingest via Abilene, we hope that a preliminary version will be available for testing by mid or late-2001.
6. **Summary and Future Directions**

Project CRAFT has demonstrated that WSR-88D base data can be delivered to virtually any user without loss of information content, and at low cost, using existing software and telecommunications infrastructures. It further has demonstrated that the long-term archival of base data at the NCDC is more effectively accomplished via internet-based transmission rather than 8 mm recording technology. The CRAFT test bed is now being used to evaluate a variety of important issues for sustained operation, such as network reliability, redundancy, and security, and efforts are underway to develop improved techniques for data compression and to explore opportunities in data mining.

The next phase of testing, known as CRAFT-2, will add some 30 WSR-88D radars to the original CRAFT framework via the Internet2 and Abilene networking infrastructures. Concurrently, LDM and data compression will be added to the Open Systems base data distribution system in an effort to facilitate the long-term viability of the CRAFT approach.

We will continue to pursue other opportunities for expanding the usefulness of CRAFT, noting that preliminary thought already has been given to including other radars, especially those operated by the Federal Aviation Administration (Weber 2000). Finally, consideration is being given to creating one or two sites, linked to Abilene and probably involving universities or NCAR, that would maintain several months of base data online for immediate perusal and download. In this regard, CRAFT represents a unique source of information for future research in data mining, digital libraries, and distributed storage technology.
7. Acknowledgments

Project CRAFT is truly a collaborative effort among numerous institutions, and the outstanding collegiality of all involved is without question the reason for its success. In that spirit, the authors gratefully acknowledge the Oklahoma State Regents for Higher Education and the National Science Foundation EPSCoR Program which together provided nearly a quarter of a million dollars of seed funding to initiate Project CRAFT in 1998. Messrs. Jason Martin and Chris Petroff, formerly with the State Regents, and James Deaton, currently with the State Regents, provided outstanding technical and administrative support, as did Drs. Richard Carpenter and Gene Bassett of CAPS. Numerous individual from the National Severe Storms Laboratory also contributed in substantial ways, including Carl Sinclair and Mark Benner. Russ Rew of the UCAR Unidata program was a key player in the use of LDM, and Woody Roberts of the NOAA Forecast Systems Laboratory provided valuable advice along the way. Guy Almes of the Abilene Program has been extremely valuable in helping take CRAFT to the next level, and Jim Belville, Walt Telesetsky and Greg Mandt of NOAA have been strong supporters of the CRAFT vision from the very beginning. The advocacy of improved access to WSR-88D data by the US Weather Research Program Science Steering Committee and, in particular, Rit Carbone and Bob Gall, is gratefully acknowledged.

This research was supported by the National Science Foundation under Grant ATM91-20009 to the Center for Analysis and Prediction of Storms at the University of Oklahoma. Additional support was provided to CAPS and the National Severe Storms Laboratory by the National Oceanic and Atmospheric Administration through Cooperative Agreement NA67RJ0150.
8. References


Rew, R. K. and G. Davis, 1990: Distributed data capture and processing in a local area network. Proceedings, 6th Int. Conf. on Interactive Information and Processing Systems for


UCAR, 1998: Unidata LDM 5 Site Managers Guide, 72 pp. [Available from the University Corporation for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.]


from School of Meteorology, University of Oklahoma, 100 East Boyd, Room 1310, Norman, Oklahoma 73019.]
Figure 1. The Operational WSR-88D Doppler Weather Radar Network.
Figure 2. NCDC archival procedure for WSR-88D Level III data.
Figure 3. Breakdown of requests for Level III data at the NCDC as of December, 1999.
Figure 4. NCDC archival procedure for WSR-88D Level II data.
Figure 5. Breakdown of requests for Level III data at the NCDC as of December, 1999.
Figure 6. Composite radar reflectivity at (a) 6:40 am CDT (1240 UTC) and (b) 10:43 am CDT (1643 UTC) on 5 April 1999.
Figure 7. (a) Model initial analysis of the 700 hPa vertical velocity field, valid at 1200 UTC on 5 April 1999, using Level III NIDS data and other observations. Vertical velocity is shaded, barbs indicate the horizontal wind, and the contours depict horizontal wind. (b) Three-hour forecast composite reflectivity, valid at 1500 UTC, based upon the initial condition shown in panel a.
Figure 8. As in Figure 7, but with Level III NIDS data replaced by Level II base data and single-Doppler velocity plus thermodynamic retrieval.
Figure 9. The Radar Interface and Data Distribution System (RIDDS) and its interface to the WSR-88D radar.
Figure 10. The Radar Interface and Data Distribution System (RIDDSS) and its interface to the WSR-88D radar.
Figure 11. The sharing of data by multiple sources from cooperating LDM systems linked to the internet.
Figure 12. Technical strategy for accessing WSR-88D base data in Project CRAFT. See the text for details.
Figure 13. The original CRAFT network as of January, 1999. The six radars, shown in black, transmit compressed base data, in real time over dedicated 56 kilobit/sec phone lines, simultaneously to ingest computers at the NSSL, University of Oklahoma, and National Climatic Data Center. The latter data travel over the commodity internet as well as the national Abilene backbone.
Traffic Analysis for Serial12/1:13
CAPS (project CRAFT) Amarillo
Weather Service field office radar - 56K

System: onenethub-okc1.onenet.net in Oklahoma
Maintainer:
Interface: Serial12/1:13 (102)
IP: No hostname defined for IP address (172.24.30.245)
Max Speed: 7000.0 Bytes/s (propPointToPointSerial)

The statistics were last updated Sunday, 13 June 1999 at 15:02,
at which time 'onenethub-okc1.onenet.net' had been up for 154 days, 10:31:18.

'Daily' Graph (5 Minute Average)

'Max In: 6350 B/s (50.7%)  Average In: 2773.0 B/s (35.6%)  Current In: 6650 B/s (9.5%)
Max Out: 2050 B/s (3.7%)  Average Out: 1260 B/s (1.8%)  Current Out: 48.0 B/s (0.7%)

'Weekly' Graph (30 Minute Average)

Max In: 6345 B/s (50.7%)  Average In: 2772.0 B/s (35.7%)  Current In: 8360 B/s (11.9%)
Max Out: 2950 B/s (3.7%)  Average Out: 1150 B/s (1.7%)  Current Out: 540 B/s (0.7%)

Figure 14. The real time network monitoring system used for the six CRAFT radars in
and around Oklahoma. Shown is a sample display for the 56K telephone line to the
Amarillo, Texas WSR-88D.
Figure 15. The real time network monitoring system used for the Hytop WSR-88D radar in Alabama. Courtesy Dr. Steve Goodman, NASA Marshall Space Flight Center.
Figure 16. *The real time data archival system now being tested at the NCDC. The commodity T1 line will become a T3 line linked to Abilene.*
Figure 17. Schematic of future numerical weather prediction models in which nested grids change automatically in response to the weather. Radar images courtesy Weather Services International Corporation.
Figure 18. University participants in Internet2 as of January, 2000.
Figure 19. a) High-speed Abilene backbone network and b) gigabit per second router points of presence (gigapops).
Figure 20.  a) Internet2 universities (yellow) and WSR-88D radar sites (black).  b) Linking each WSR-88D to the nearest Internet2 node using 56K phone lines.
Figure 21. Logical topology for ingesting WSR-88D base data using the Internet2 and Abilene infrastructures. Data from the each is transmitted via a 56K phone circuit (red line) to the nearest Internet2 university (red dots), each of which is linked to Abilene (violet circle). Data from all radars is transmitted around the Abilene ring by certain institutions that operate Abilene gigabit routers (gigapops) and that have high quality of service needs (e.g., NCEP, NCDC). Other users (aquamarine oval) can obtain data via the commodity internet, which links to Abilene, or from Abilene directly.
Figure 22. CRAFT-2, an extension of Project CRAFT to additional radars nationwide using the Abilene and Internet2 infrastructures.