FINAL PROJECT REPORT
To
UCAR / COMET Partners Program.

December, 2010.

Project Title: *Implementation of Hydrologic Information Data Server (HIS) in the Dallas/Fort Worth Metroplex.*

University: University of Texas at Arlington, Department of Civil Engineering, 405 Nedderman Hall, Arlington, TX 76019

Name of University Researcher Preparing Report: John A. McEnery

NWS Office: West Gulf River Forecast Center, 3401 Northern Cross Boulevard, Fort Worth, TX 76137

Name of NWS Researcher Preparing Report: Paul W. McKee

UCAR Award No.: Subaward No. S09-81073

SECTION 1: SUMMARY OF PROJECT OBJECTIVES.

The object of this project was to establish a data platform on the campus of the University of Texas at Arlington configured to store and host hydrologic information from the National Weather Service - West Gulf River Forecast Center (WGRFC) for the Dallas / Fort Worth area. The original proposal describes a one-year pilot collaboration between UTA and the NWS-WGRFC to tailor a gateway data system for ingesting and disseminating WGRFC daily precipitation data. That data is maintained and archived in a standard format using Microsoft SQL Server. From this pilot project inter-agency collaboration is continuing for expansion of the data server to host additional hydrologic data sets of specific value in the WGRFC service area and of national value.

In its broadest sense, the goal of this project was to share regional hydrologic information derived from the NWS WGRFC among a community of stakeholders; including researchers, forecasters, students, educators and regional or local experts. The objective components were to assess data/database format, design methodology for transfer to archive structure, populate server, verify data, deploy and maintain. These objectives have been met and exceeded.

The project between UTA and the NWS established a common data link for managing and serving the NWS precipitation data through a university hosted server which is shared via the CUAHSI HIS. Plans are developing to utilize this system for hydrologic data sharing across multiple government agencies with a primary goal of supporting water management decisions at the Tarrant Regional Water District (TRWD). Through a multi-agency, collaborative effort,

which includes the U.S. Geological Survey, the U.S. Army Corps of Engineers, and Tarrant Regional Water District, the project is now expanding. Researchers and students across the partner agencies are coordinating to develop procedures for managing and serving additional hydrologic information. A long term goal is to host a suite of hydrologic data sets of local, regional, and national importance. Ultimately, this increased access to data will enable the hydrologic community to develop new applications, products, and services for stakeholders. In addition, the NWS is advocating this pilot project become a testbed for the Integrated Water Resources Science and Services (IWRSS) initiative. IWRSS is a multi-year federal initiative to streamline the accessibility and integration of hydrologic data and forecasts among government agencies for both private and public interests.

SECTION 2: PROJECT ACCOMPLISHMENTS AND FINDINGS.

Describe the research/development activities and accomplishments carried out to date. These accomplishments may relate specifically to the original project objectives, or they may be ones that arose during the course of the project (e.g., development of an innovative method for accomplishing the objective or insight into a related problem). Highlight any major changes to the scope of work. If the project involved separate research topics, please list each separately

TECHNICAL OVERVIEW

The NWS West Gulf River Forecast Center (WGRFC) generates Multi-Sensor Precipitation Estimates (MPE) through a multi-stage process; blending information from Nexrad WSR-88D (Weather Surveillance Radar 88 Doppler), physical gages, community collaborative reports and satellite observations. The MPE values are estimated at locations assigned on a grid coordinate system referred to as the Hydrologic Rainfall Analysis Project (HRAP) having approximately a 4 km by 4 km spacing. The MPE product is originally created in an XMRG file format. The extent of WGRFC coverage is shown in the following Figure 1. The closely spaced red icons each represent the center of an HRAP grid cell, displayed in the CUAHSI HydroDesktop program.

The WGRFC archives MPE product in a reduced text form that is derived from the XMRG files. These CSV files list HRAP ID number, HRAP X &Y (analogous to latitude and longitude) and MPE values. Each file contains 169,000+/- rows corresponding to each of the 169,000+/- HRAP locations within the WGRFC service area. Time (in UTC format) is identified as part of the file name for each CSV file.

This project has followed protocols developed by the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI). CUAHSI is a group of more than one hundred affiliated universities and research organizations with the shared goal of advancing research and education in hydrologic science. CUAHSI provides oversight for the formation of a network of Hydrologic Information Servers (HIS). The HIS system is a framework established by CUAHSI for the exchange of hydrologic data in a common format know as WaterML (Water Markup Language). The servers each provide access to information from its respective data source(s). WaterML provides a common archival language for the exchange of data via all servers in the CUAHSI system. By using this common language, data can be fluently exchanged regardless of whether it originated in the EPA's STORET format or the NWS SHEF format, etc

Daily dataset within U.S. & Mexico

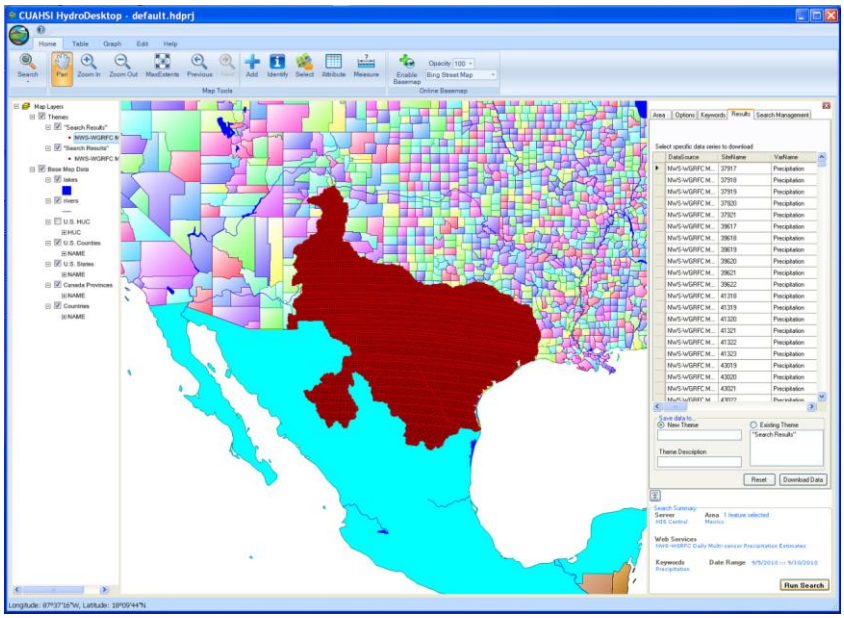


Figure 1. Coverage of daily MPE data within West Gulf River Forecast Center service area.

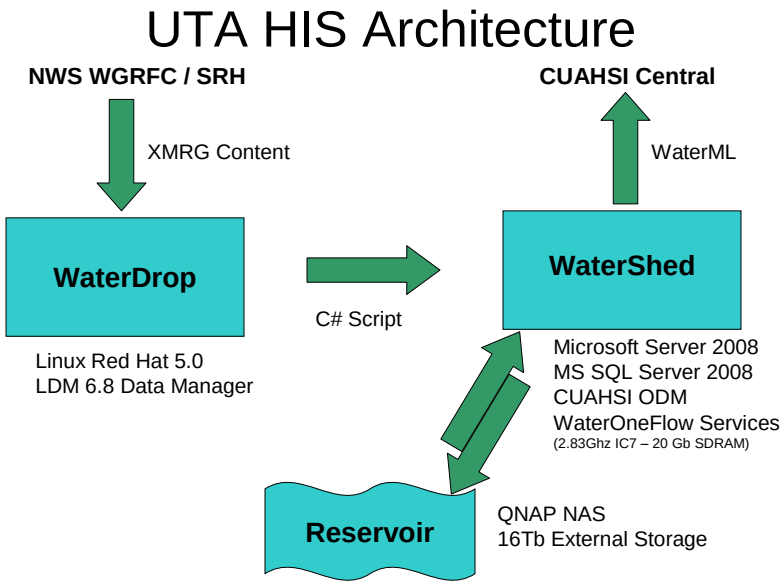
The first task combining the efforts of UTA and the NWS-WGRFC was to assess the data format for the precipitation estimates (MPE) that the UTA server handles. That addressed the existing format of the NWS data in comparison to the WaterML format and transfer methods. Task 1 also addressed volume requirements for mass data storage. Due to security issues, data size and format concerns, it was unclear after the initial meeting how difficult it would be to handle the data. Discussions between the collaborators led to the conclusion that it would be feasible to ingest all of the data specified in the original proposal, and more. The objective stated in the original proposal indicated that dissemination of daily precipitation values for DFW area would be accomplished. Based upon what was learned, the objectives were broadened; both spatially and temporally. Hourly (as well has daily) values are being hosted with coverage to include the entire WGRFC forecast area. This broader spatial extent covers most of Texas, as well as parts of Mexico, New Mexico, and Colorado (Figure 1 above). The existing archived data sets were prepared by NWS personnel and transferred to UTA on a single portable hard drive. There were 2 archives; one containing hourly values beginning in 1997 and another containing daily values beginning in 1995.

The second task was to populate the server with the daily precipitation data and maintaining transfer of updates. The precipitation data that was transferred in the archived files exist in a format different from the WaterML structure native to the CUAHSI data server. Reordering that data and placing it in the Operational Database Model (ODM) was a relatively simple task. However, the sheer number of data lines involved (millions) made the task daunting. Therefore, an assimilation program was written in the C# programming language to accomplish the data transformation and loading task. That programming produced a data loader that has been a successful and useful tool for data population. After archived values of

daily MPE were ingested, the next effort was to populate the server with the archived hourly values. This task was complete during July 2010.

Entering these values into the ODM required writing a custom dataloader. The standard CUAHSI dataloader was developed to read values at a small number of sites through a period of time. The data that has been ingested from WGRFC represent a very large number of locations, all at a single moment in time. The UTA dataloader was written in C #. An additional function that the UTA dataloader had to accomplish was to read the UTC from each file name and assign that along with the other values that were ingested into the SQL database for each HRAP center. The ODM requires both local and UTC time, so the UTA dataloader was designed to convert between the 2 values while dynamically adjusting for the annual onset and offset of daylight savings time and 2 different time zones.

The combined storage size of these data sets when saved within the SQL database is approximately 2 Tb. Three pieces of equipment are being used. A schematic of this architecture follows.



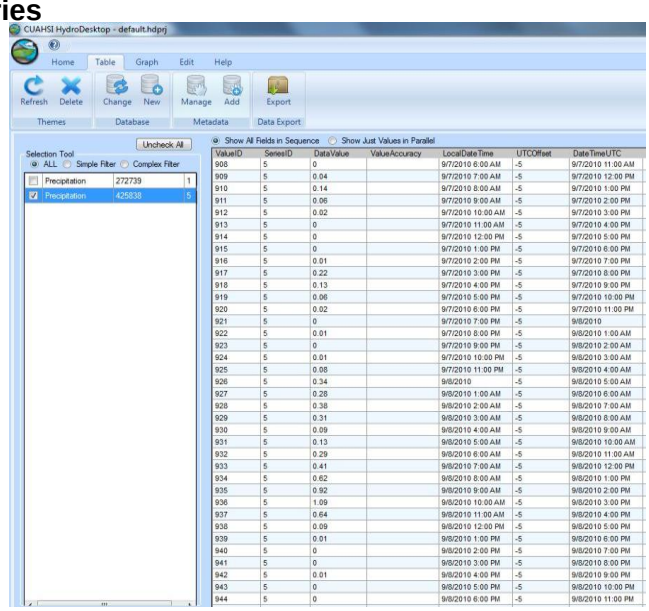
1. (Watershed) The primary workstation is a Dell Precision with an Intel IC7 2.83 GHz quad core processor. Memory has been expanded to 20 Gb and 2 additional high-speed Ethernet slots have been added. The operating system is Windows Server 2008 and the database is SQL Server 2008 (32-bit with AWE enabled).

2. (Reservoir) The storage device is a product from QNAP, Inc. It is a network attached storage device (NAS) which houses 8 2Tb enterprise class Seagate drives for a total storage capacity of 16 Tb. The NAS has onboard 2 Gb memory and its own Intel Core 2 Duo 2.4 GHz processor. It is directly attached to the workstation by Ethernet. The drive configuration is RAID 5. The SQL Server is configured to recognize the storage device as an iSCSI target.

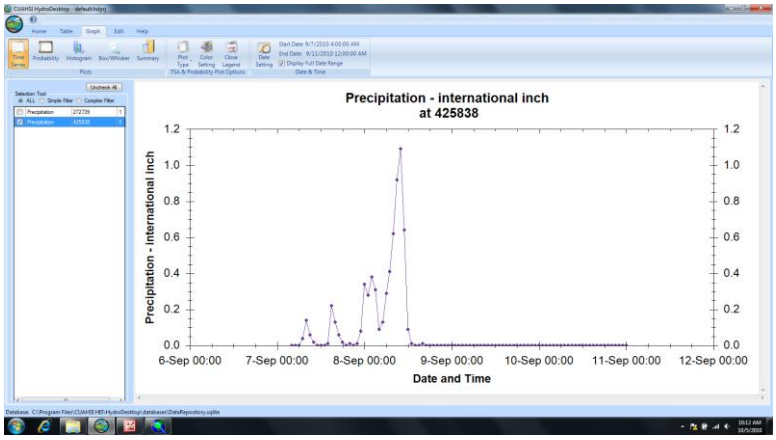
3. (Waterdrop) The secondary computer is a standard Dell desktop formatted with the Red Hat Linux 5.1 operating system. 2 additional high-speed Ethernet slots have been added. Unidata Local Data Manager (LDM) 6.8.1 is installed and will be used to receive hourly updated CSV files from NWS Southern Region Headquarters (SRH). The folder to which LDM saves data is shared and mapped on the primary workstation (Watershed) as a network drive. Waterdrop and Watershed are directed connected by an Ethernet line. The UTA data loader (which runs on Watershed) will read the latest data file from Waterdrop each hour and ingest that to the SQL database, keeping the ODM continuously updated. FTP has also been installed on Waterdrop to allow external access to the raw CSV files.

The CUAHSI WaterOneFlow vers. 1.1 services have been installed on Watershed and registered to publicly expose data. The largest dataset currently registered is the daily set of values covering the entire WGRFC service area. Smaller coverage sets of hourly data are also registered. Efforts continue to broaden the scale of hourly coverage in space and time, and speed up download performance. The resulting public access to hourly MPE data is displayed in the following two images. The first is a tabular listing of recent values specific to an HRAP center in Tarrant County, Texas (Fort Worth/Arlington) during the Tropical Storm Hermine September, 2010. The second figure is a graph of that precipitation estimate series.

Tropical Storm Hermine
Hourly Time Series
Arlington, TX



Tropical Storm Hermine
Hourly Time Series — Arlington, TX



Through the remainder of the project timeline, efforts focused upon steps in three key areas bringing the project to completion. The team registered multiple datasets through a publicly accessible catalog for discovery and download. Optimization of the SQL databases through the development of a database index strategy has resulted in reliable download functionality. The WCRFC, NWS Southern Region Headquarters and UTA implemented an internet based connection using the Unidata Local Data Manager (LDM) to transfer current updates to UTA. Using that connection, CSV files from NWS Southern Region Headquarters (SRH) with current updates are received each hour for entry into the database and maintaining a current information source.

In addition to the scope of work defined within the COMET proposal, the area of coverage and frequency of coverage has been expanded to hourly MPE across the entire WGRFC forecast area. The first complete dataset registered was the daily MPE series for the entire WGRFC. Hourly coverage for several areas of the RFC have been completed and developments continues to complete an hourly set for WGRFC wide coverage.

One of the problems encountered was the scalability consideration associated with the development of very large SQL datasets. Very long data retrieval times became inoperable. To overcome that problem database features known as database search indexes were generated. These constructs reside within the SQL database and are useful for data retrieval. They were written and applied to several of the larger data sets. The positive finding was that the indexes substantially reduced data retrieval time. Data download times for registered sets of Daily and Hourly values now range between quick to nearly instantaneous! However, they presented a performance trade off. The negative outcome was that the search indexes substantially slowed the data entry process for adding more data. Nearly an hour to load a one hour set of values. Not a sustainable situation. The solution that has been developed is to maintain two separate databases; recent values versus archive values. Recent values are entered into a small dynamic set containing a short period of record (no indexes needed). This receives current updates as they arrive each day and hour. Then the second database is a large static set with the full 15/17

yr history of archive values. This set will only have data added quarterly or semi-annually after several months of values have accumulated in the smaller current values set.

Our combined resources and connections have sparked a renewed interest with the Texas Natural Resources Information System (TNRIS) to utilize their resources to serve precipitation data for all of Texas. TNRIS is a component of the Texas Water Development Board and is the state's clearinghouse and referral center for Texas natural resources data and information. The WGRFC and UTA have formed an agreed with TNRIS adapting this project for State-wide application so that to this system will serve as the deliver vehicle for MPE data from WGRFC to TNRIS.

SECTION 3: BENEFITS AND LESSONS LEARNED: OPERATIONAL PARTNER PERSPECTIVE.

List the benefits to the NWS office from the collaboration and any significant lessons learned during the study. Please be as specific as possible, particularly in regard to any improvements in forecasting resulting from the COMET project (see examples). Identify any major problems encountered and describe their resolution.

Benefits of collaboration.

The benefits to the NWS and WGRFC from collaboration through this project have been extensive and continue to deepen. Through the course of this project, numerous meetings and conference calls have strengthened the relationship between NWS and UTA setting stage for continued, more frequent collaborative research. In addition, this project has enhanced and increased collaboration with other agencies as well. Presentations and meetings with other state and federal agencies, as well as, river authorities and water districts to discuss ideas and share project objectives and successes is beginning to birth other projects and build national attention. These agencies include the US Geological Survey, US Army Corps of Engineers, Texas Water Development Board(TWDB and associated Texas Natural Resource Information System), Tarrant Regional Water District (TRWD), Lower Colorado River Authority, and Brazos River Authority. This COMET partner project has become the catalyst for another project funded by TRWD to establish a common data sharing repository for federal water data sources needed to force a newly developed real-time floodway model. In addition, extension of this project is underway to include precipitation data sets from two neighboring NWS River Forecast Center domains (Arkansas Basin RFC and Lower Mississippi RFC), initially to complete the entire state of Texas at the request of TWDB for emergency management decision support, but also as a first step toward national extension to serve precipitation data publicly which will continue to promote the NWS mission and UTA acclaim across the country.

Improved collaboration with UTA and partner agencies is leading toward the development of a common data warehouse for water agencies in North Texas. As a result, streamlined exchange of data will improve the efficiency of the forecasting process in support of the NWS mission. This project will become the basis for standardization and protocol for other projects funded to extend coverage to other parts of the nation. As NWS continues long-standing discussions about how to effectively and efficiently archive and serve its own precipitation products, protocols and standards from this project will facilitate and direct NWS internal discussions and planning to accomplish this service gap. In addition, technologies transferred from this project will ensure that necessary links are maintained such that any data

served by NWS can be easily connected to and accessible by the CUASHI HIS. This paradigm could interest other universities to extend similar projects for surrounding regional areas.

Lessons Learned.

Many lessons were learned through resolution of problems encountered along the way through the progress of this project. WGRFC learned that text file formats of the native XMRG format is a better way to share, transfer, and archive data than GIS shapefile format. This effectively hardened the WGRFC archive of precipitation data sets and made the process of filling gridded precipitation data requests more efficient. The transfer of data using LDM strengthened the NWS support and knowledge of LDM process and scripts required to securely transfer data real-time between NWS and UTA. This knowledge and experience is invaluable to support the ongoing transition of the NWS away from FTP to LDM to handle all data transfer with non-NWS sources. Finally, the increased awareness and capability of CUASHI HIS and its paradigm of data access and sharing that continues to be investigated to support the multi-federal agency IWRSS initiative. This awareness should ultimately prompt and facilitate other NWS products to become available through a CUASHI HIS (i.e. river forecasts, COOP observations, and a host of weather parameters, etc.)

SECTION 4: BENEFITS AND LESSONS LEARNED: UNIVERSITY PARTNER PERSPECTICVE.

Describe the benefits to the University resulting from the collaboration and any significant lessons learned during the study. Identify any major problems encountered and describe their resolution.

This project is yielding benefits beyond just the HIS system. The student participant is gaining experience with several CUAHSI hydrologic data tools, GIS, programming functions and the relationship between these tools with Microsoft SQL. He has learned about the forecast operations and products of the NWS. The project has served as a pilot for future coalescing of data in a regional repository with participants from additional agencies. As the system continues to grow into a truly multi-agency platform, it will be an increasingly valuable information resource for hydrologic information on the UTA campus and elsewhere. An additional extension of the work has already been generated partnering the NWS, UTA and the Tarrant Regional Water District. The new project will extend the COMET pilot project with the inclusion of data from more sources and adapt this system to provide inputs for a regional floodway model.

This project has helped to build a relationship between WGRFC and UTA that will facilitate future research collaboration, possible student internships, and shared lectures and seminars. UTA anticipates continued collaboration with WGRFC as research needs and funding sources allow.

SECTION 5: PRESENTATIONS AND PUBLICATIONS

Provide complete citations using the AMS bibliographic format for each thesis, dissertation, publication or presentation prepared as part of this project.

No publications have yet been developed from this project. However, a graduate student at UTA has devoted his degree work to this project. As he nears the completion of his graduate studies it is anticipated that he will generated a set of documents. These are expected to include a journal paper for publication in a peer reviewed research journal and a manual providing detailed documentation of all steps in the implementation and operation of the system. This will be suitable for use by other RFCs that may wish to partner with Universityies to host MPE data. Combined these will constitute a graduate thesis for partial fulfillment of a Master of Science degree. The likely venue for journal publication will be the Bulletin of the AMS.

Several conference presentations of this project have already been given and others are currently scheduled. Completed conference presentations are as follow;

McEnery, J., McKee, P., Shelton, G., *Implementation of a Hydrologic Information Data Server in the DFW Metroplex.* American Geophysical Union Fall Conference 2010: San Francisco, CA, Dec. 16, 2010. (Poster)

McKee, P., McEnery, J., Shelton, G*., Implementation of a Hydrologic Information Data Server in the DFW Metroplex.* Texas Floodplain Managers Association Fall Conference 2010: San Marcos, TX, Nov. 8, 2010.

McEnery, J., Invited Speaker: *Implementation of HIS for U.S. National Weather Service.* Texas Flood Management Conference: University of Texas, Austin, TX. Oct. 8, 2010.

McEnery, J. Invited Speaker: *Hydrologic Information System for Dallas/Fort Worth.* Texas FloodPlain Managers Quarterly Section Meeting: UTA, Arlington, Texas. Sept. 2, 2010

McEnery, J. Invited Speaker: *Regional Hydrologic Information System for North Texas.* Tarrant Regional Water District Interagency Quarterly Roundtable Seminar. Fort Worth, TX. March 31, 2010

SECTION 6: SUMMARY OF UNIVERSITY/OPERATIONAL PARTNER INTERACTIONS AND ROLES.
Describe the responsibilities of the various project participants over the course of the entire project.

Collaboration between UTA and the NWS has been essential to fulfilling the project goals. There has been an ongoing sequence of meetings and conference calls between the partners. UTA and WGRFC have shared equally in assessment of the data management problem. The WGRFC has provided knowledge of the existing data format and was responsible for preparing transfer files. UTA has handled the assimilation of that data into the HIS database, development of the system architecture and the registration of the system for public access.

Originally, it was unclear how to approach the practical problem of managing and transfer large volumes of archived data. The problems have included developing scripts to recognize the format of the available data, write that into a structured database, tune the performance of that database and implement a system that will automatically update the

database with new MPE values. All of these steps required coordinate with the WGRFC, as well as the dedicated support from their personnel. Through this project, the process of working together has produced a much better understanding of the WGRFC as an organization and a skilled set of individuals. WGRFC provided background, overview, and explanation of MPE data sets. WGRFC also provided support for establishing the LDM data transfer connection. UTA and WGRFC have also both shared in the effort to make presentations of the work and to purse avenues for extention.