Computers & Geosciences 50 (2013) 145–153

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Hydrologic information server for benchmark precipitation dataset

John A. McEnery[a,*], Paul W. McKee[b,1], Gregory P. Shelton[b], Ryan W. Ramsey[c]

[a] Department of Civil Engineering, University of Texas at Arlington, 405 Nedderman Hall, Arlington, TX 76019, USA
[b] US National Weather Service, West Gulf River Forecast Center, 3401 Northern Cross Boulevard, Fort Worth, TX 76137, USA
[c] Department of Civil Engineering, University of Texas at Arlington, Nedderman Hall, Arlington, TX 76019, USA

ARTICLE INFO

Article history:
Received 12 April 2012
Received in revised form
31 July 2012
Accepted 1 August 2012
Available online 16 August 2012

Keywords:
NEXRAD
MPE
Precipitation
WaterML
CUAHSI

ABSTRACT

This paper will present the methodology and overall system development by which a benchmark dataset of precipitation information has been made available. Rainfall is the primary driver of the hydrologic cycle. High quality precipitation data is vital for hydrologic models, hydrometeorologic studies and climate analysis,and hydrologic time series observations are important to many water resources applications. Over the past two decades, with the advent of NEXRAD radar, science to measure and record rainfall has improved dramatically. However, much existing data has not been readily available for public access or transferable among the agricultural, engineering and scientific communities. This project takes advantage of the existing CUAHSI Hydrologic Information System ODM model and tools to bridge the gap between data storage and data access, providing an accepted standard interface for internet access to the largest time-series dataset of NEXRAD precipitation data ever assembled. This research effort has produced an operational data system to ingest, transform, load and then serve one of most important hydrologic variable sets.

1. Introduction

1.1. Objective

The National Weather Service (NWS) generates Multi-Sensor Precipitation Estimates (MPE) through a multi-stage process. Weather radars can provide high-resolution precipitation data in space and time. These rainfall estimates, based upon reflectivity from NEXt generation weather RADar (NEXRAD), are then refined using observations from physical gauges, community collaborative reports and satellite sensing. The multisensor precipitation estimator process and the programs used to derive these multisensor fields are described in Lawrence et al., (2003). Fig. 1 shows the mosaic of NEXRAD radars utilized by the NWS West Gulf River Forecast Center (WGRFC). The resulting MPE values are estimated each hour and issued both as hourly and compiled daily products Generation of MPE has been ongoing since the mid-1990s. MPE is spatially distributed on a grid, having approximately a 4 km by 4 km spacing that is referred to as the Hydrologic Rainfall Analysis Project (HRAP) grid coordinate system. The grid is based on a polar stereographic map projection with standard latitude of 60° North and longitude of 105° West (NWS, 2001). This MPE data is considered by the NWS to be the "best estimate" of rainfall. It is a valuable hydrologic data resource, providing spatially continuous estimates at small time spacing to bridge the temporal and spatial gaps that existed in the past when relying solely on physical gauges.

The data has been stored, however, in agency specific formats without an efficient means for public consumption. The MPE product is originally generated in a binary format called XMRG. Depending upon the respective River Forecast Center, subsequent binary forms of the data may also be saved either as shapefiles or as netCDF files generated using the version 3.5.1 library routines (Rew and Davis (1990)) Availability of these alternate formats is not uniform across the 13 River Forecast Centers. Applications such as hydrologic models require rainfall information in a sequential history specific to the region of study. Therefore, where data is saved in binary format as files for individual times, sets must be compiled and then parsed to extract a sequential series of information at a specific location or sub-region. Consequently, over its 17 year history, use of MPE among the scientific and engineering community has been very limited. The objective of efforts presented in this research paper is to provide previously unavailable access in time-series format to an extensive volume of NEXRAD MPE information covering the WGRFC service area (1 of 13 in the United States) through a proven service-oriented architecture.

1.2. Pertinent developments

Many collections of observational data for the geosciences now exist through internet accessible services. When working with

* Corresponding author. Tel.: +817 272 0234.
E-mail addresses: mcenery@uta.edu (J.A. McEnery), paul.mckee@noaa.gov (P.W. McKee), greg.shelton@noaa.gov (G.P. Shelton).
[1] Tel.: +817 831 3289.

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.08.005

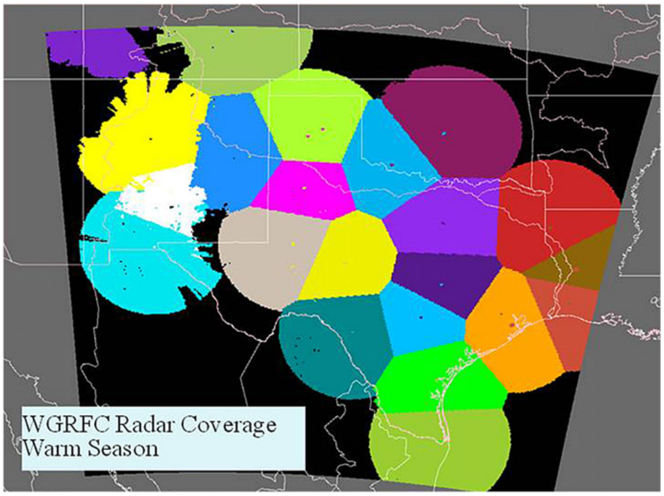

Fig. 1. Mosaic of NEXRAD radars utilized by the NWS WGRFC (NWS).

data from numerous agencies and sources, there are often variations between the terminologies used to describe observations as well as the methods for managing and communicating data. (Horsburgh et al., 2009; Sheth and Larson, 1990). Greater efficiency and interoperability is achieved through the communication of geosciences data in a standardized machine accessible protocol (Goodall et al., 2008). The use of standardized eXtensible Markup Languages has emerged as a means for pass data from computer to computer over the internet as XML documents. Among these is the Water Markup Language (WaterML) (Zaslavsky et al., 2007).

The Consortium of Universities for the Advancement of Hydrologic Science, Inc., (CUAHSI) has developed methods and applications for ingesting and disseminating hydrologic data in a service-oriented architecture (SOA) framework (Maidment 2008a,b). The CUAHSI Hydrologic Information System (HIS) uses a layered map-based ontology, developed using Web Ontology Language (OWL) (Piasecki and Beran, 2009). The operational structure of the HIS service-oriented architecture consists of three components with complementary functions. The first of these is the HydroServer, which is a suite of software products used for data storage and publication. The second is HIS Central, which is a central index that contains cataloged metadata describing the data contents of all HydroServer services registered with HIS Central. It functions as searchable index for the discovery of published data series based upon metadata characteristics such as site location, variable and temporal characteristics. While it responds to client queries with the descriptors of data series characteristics, accessible address, etc., it does not store the actual data values specific to each of the registered HydroServers (Whitenack, 2010). Third are the data access applications which include HydroDesktop and other clients used to search, download and view hydrologic data.

One of the features of the HydroServer suite is the data model specifically structured for the storage and retrieval of hydrologic observations. This is referred to as the Observations Data Model (ODM) (Tarboton et al., 2008; Horsburgh et al., 2008). ODM is a schema created as a generic template for use with a relational database. One of the major challenges of the hydrologic data infrastructure is elimination of the data ambiguities and heterogeneities that arise from differences between various data collectors and investigators (Maidment, 2008a,b). The ODM is designed to help overcome these data disparities by creating a standardized format for storage of point observations, along with sufficient metadata, or ancillary information corresponding to the actual data values.

The ODM consists of a series of tables within a relational database format, along with a significant amount of ancillary data that ultimately describes the individual point observations that are stored as records or rows in the database. The ODM structure contains the following 27 tables; Categories, CensorCodeCV, DataTypeCV, DataValues, DerivedFrom, GeneralCategoryCV, GroupDescriptions, Groups, ISOMetadata, LabMethods, Methods, ODM Version, Offsets, Qualifiers, QualityControlLevels, SampleMediumCV, Samples, SampleTypeCV, SeriesCatalog, Sites, SpatialReferences, SpeciationCV, TopicCategoryCV, Units, ValueTypeCV, VariableNameCV, Variables and VerticalDatumCV. The table names with the suffix CV are those that define controlled vocabulary. The DataValues Table is one of the most primary; containing 17 fields which reference elements such as the numeric value of the observation, date and time at which the value was observed and the integer identifier for the variable that was observed. The SeriesCatalog Table contains 30 fields which reference some elements also appearing in the DataValues Table and other elements such as variable type, series begin time and series end time. Therefore it must be updated each time a new hourly set of values is inserted in the database. The ODM is a generic template, not restricted to any specific computer software. CUAHSI has also created and provided a Microsoft SQL Server blank database schema that can be downloaded and attached as a database within the SQL Server Management Studio. The ODM was originally described in 2005, and the most current edition is ODM Release Version 1.1.

Another feature of the HydroServer suite is the component called WaterOneFlow (WOF) Web Services. WaterOneFlow is an application which is installed along with the ODM on a host computer. A separate instance of WaterOneFlow is required for each ODM on a host computer. This application is the utility that responds to queries from clients seeking specific data. Upon receipt of a client query, WaterOneFlow searches the associated ODM to retrieve the requested information and then returns that data as an XML document in the language described above WaterML. The WaterOneFlow search functions for retrieving ODM information include GetSiteInfo, GetSiteInfoObject, GetSitesXml, GetSites, GetVariableInfo, GetVariableInfoObject, GetValues and GetValuesObject (Beran et al., 2009; Piasecki and Beran, 2009). The values returned to a client as a WaterML document are useful as time-series information. WaterOneFlow provides a machine-readable description of the operations offered by the service written in the Web Services Description Language (WSDL) (Skonnard, 2003). WaterOneFlow exposes both Simple Object Access Protocol (SOAP) and Representational state transfer (REST) compliant endpoints.

2. Methodology

2.1. Project approach and system hardware

To address the limitations of the Multi-sensor Precipitation Estimate data access described in the introduction, the project investigators and personnel from the NWS WGRFC have partnered, tailoring a platform to store and serve MPE data through a university hosted computer. These MPE values are ingested into a relational database and disseminated through an accessible web service. The methodology followed draws upon many of the SOA developments of CUAHSI HIS.

An alternative strategy to deliver MPE data could have been pursued using a services stack based upon the Thematic Realtime Environmental Distributed Data Services (THREDDS) (Domenico et al., 2002). THREDDS Data Server (TDS) is a web server with functionality to subset data spatially and temporally from larger datasets. Datasets are maintained in file catalogs and netCDF is one of the supported file formats that can be read by TDS. Deploying TDS with an OGC compliant service such as Web Feature Services (WFS) might also accomplish the delivery of MPE. However, this would require additional development and would be subject to varying availability of netCDF formatted MPE product across the 13 NWS River Forecast Centers. Because there is already an established research community using WaterML and a suite of tools available to deliver and consumer data in WaterML, the objective of this research has been focused solely upon the delivery of this valuable dataset using WaterML and WaterOneFlow services from CUAHSI.

The operational location for this effort is a data center in Fort Worth, Texas conforming to standards for a Tier III facility as defined by the ANSI/TIA-942, Telecommunications Infrastructure Standard for Data Centers. The dedicated primary computer (HydroDB) is a Dell PowerEdge R710 with dual 6 core hyper-threaded Intel(R) Xeon(R) processors at 3.37 GHz and 96 Gb installed memory (RAM). The computer has two separate hard drive arrays. The first drive array holds 2 Tb of storage for OS, log and system files in a RAID 50 configuration. The second drive array holds 20 Tb of storage for project data in a RAID 5 configuration. An additional computer (HydroApp) is a virtual machine configured with VMWare on the facility's central computing cluster.

2.2. Source data

The native XMRG format for MPE data is cumbersome for direct manipulation. Therefore, the WGRFC has written a read/write program named REPO to convert the unconventional XMRG file format into a more manageable text file format. REPO works as a summation program that adds the hourly XMRG files and combines them into the more familiar text file format that is being received in this project (Thompson, 2010). It reads command line arguments (input XRMG files), creates a virtual XRMG in dynamic memory that begins with zero values, and then reads the actual XMRG values, adding these to the created virtual one. After reading all the input XMRG values/files and summing up the values in the virtual memory, the program then writes these combined values based on site location into a readable row/column text file format (Thompson, 2010).

The REPO program is executed by the WGRFC, as a standard processing step when each new MPE set is generated. The XMRG output is translated into the comma-separated (csv) text file "REPO" formatted product containing each HRAP grid cell's MPE value for a given time period. Each text file contains three columns of information. The first is a column titled "lid" which represented a list or unique sequencing number for the file and had no significant meaning to the data itself. The second column is "cid," representing the conus HRAP id which is the unique site identification value for each HRAP grid cell. This value became the primary naming convention for each site within the ODM. The final column in each text file is the MPE value itself, named "valin," representing the MPE value recorded in inches.

To receive current sets of the MPE product files generated by the WGRFC, a continuous file transfer process was established between the WGRFC offices and the HIS system located at the Tier III data facility. This was achieved using the Local Data Manager (LDM) developed by the Unidata Program Center, a member of the University Corporation for Atmospheric Research (UCAR). The Unidata developed LDM, a file transfer software commonly used by the NWS which is publicly available for download from http://www.unidata.ucar.edu/software/ldm/. LDM software treats data as an opaque unit and is thus capable of relaying most types of data. The data is acquired by the software and then shared to networked computers (Unidata, 2011).

LDM was designed to run on UNIX/Linux workstations; therefore, a separate Linux based computer (virtualized with VMWare) is used for this project to accomplish LDM transfers. File transfers are initiated from the WGRFC and are routed though a central server at NWS Southern Region Headquarters and relayed to the Linux computer, HydroApp at the data facility.

2.3. System architecture

The operating platform for maintaining the MPE dataset draws upon the CUAHSI HIS developments for SOA cited above. The primary computer, HydroDB, runs the Microsoft Windows Server 2008 R2 operating system and functions as a HydroServer using the CUAHSI HIS applications ODM and WaterOneFlow Services. Data files transferred hourly from WGRFC are saved into folders on the Linux computer, HydroAPP. The creation of an active directory environment on the Linux computer was accomplished by using an open source program Samba (http://www.samba.org/samba/). Samba has been installed on the Linux computer to enable interoperability between it and the Microsoft computer. Using the TCP/IP protocol, Samba enables the client Microsoft computer to recognize and access the folders on the Linux to which the MPE files have been stored by LDM. The folders receiving MPE files on HydroAPP are essentially mapped to the HydroDB server in a manner similar to the mapping of folders shared on networked Microsoft computers. A diagram of the data flow across this platform is presented in Fig. 2. Details of the interrelated system components and the steps to implement them are described in the following sections.

2.4. MPE server

The MPE dataset is maintained and archived in a Microsoft SQL 2008 Server (HydroDB). However, to achieve the flow illustrated in Fig. 2, due to the uniqueness of the NWS data and the large volume of the MPE data stored, several procedures supporting the operational processes had to be developed and tailored. This section explains the steps that were required to translate the original MPE values into a format that could be readily inserted into the database. The first of these was application of the database schema and consequently, adaptation of the ODM from CUAHSI HIS.

Over the NWS HRAP coordinate grid, estimated MPE values are associated spatially with the center of each respective grid cell. These MPE point locations were translated into sites defined by latitude and longitude to compile the necessary ODM "sites file" descriptors. The data contained in the REPO files also had to be filtered to eliminate many MPE values located beyond the actual forecast boundaries of the WGRFC. The site locations where MPE values are recorded have been defined in an. ESRI shapefile provided by the WGRFC. The shapefile is geo-referenced in the North American Datum of 1983 (NAD83) coordinate system. Although there are a total 165,750 points that appear as HRAP centers in the shapefile and each corresponding REPO formatted data file, only 69,830 of these HRAP center points are actually contained within the boundaries of the WGRFC service area. From these shapefiles, the fields for ID, HRAPX, HRAPY, LAT, and LON were extracted and filtered to populate the ODM sites table. The ID is the numeric identifier for the center of each HRAP cell. Spatial reference for each site ID in the database is provided in both the NAD83 and HRAP coordinate systems. HRAP was not one of the spatial references original existing in the CUAHSI Master Controlled Vocabulary Registry for ODM 1.1., but it has been submitted and added to the Variables Table, as well as a

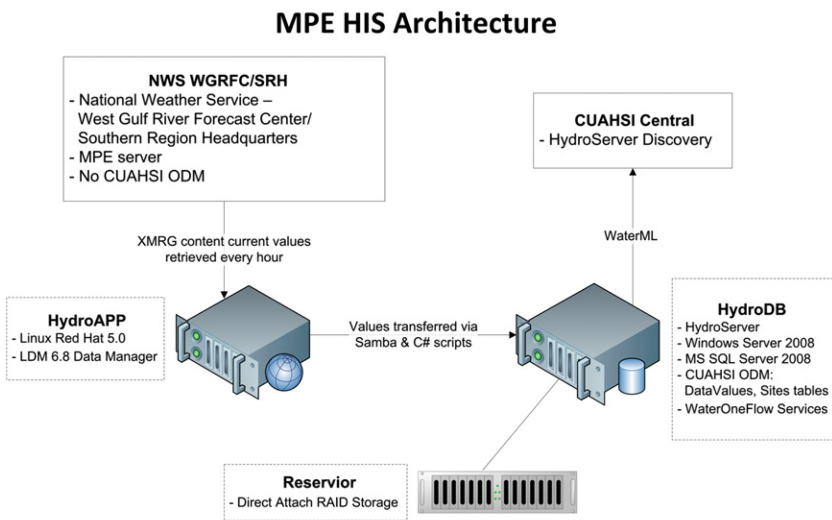

Fig. 2. Flow diagram of MPE data server.

description of the field and how the information was determined for the MPE data.

2.5. Data loader

Due to the format and volume of MPE values contained in each respective REPO file for each hourly time period, it was determined that a custom written utility was needed to automate the loading of values into the database. The processing required to write these values into a relational database has a greater performance cost than the cataloging of gridded data in flat file format. However, mapping this data to the relational database schema allows access to this information using a detailed ontology shared among a large research community. Microsoft Visual C# was selected to accomplish the programming of a multi-threaded code aptly named the MPE Data Load program (MPE_DL). This was designed to execute a number of tasks, including reading files from the mapped folder on the Linux computer and loading MPE values into the ODM database. The time stamp for each set of values was written as part of the REPO file name, so upon reading the file path argument, MPE_DL inserted the end of the time interval as the DateTimeUTC value in the DataValues table. In order to convert this DateTimeUTC to the LocalDateTime, the program referenced a "Sites Dictionary" that was created to pre-define the time zone in which each site exists, based upon grouping in one of three two-comma separated files previously. Then, after the after the time zone of each site is defined, a built-in C# function converts the original UTC time value to the local time value. MPE_DL repetitively performs the tasks of reading each MPE text file, inserting the appropriate fields for the SiteID and DataValue tables and then updating the Series Catalog Table. The flow chart in Fig. 3 shows the overall procedures of the MPE data loading program. In addition to the creation of XMRG/REPO files for new hourly time periods, the WGRFC also may conduct a post product review and release a revision of a previous file. These revisions are also received on HydroApp via LDM. Therefore, MPE_DL has been designed to recognize the differences in revised data files and, if appropriate, to rewrite the entire new value set to the Data Table.

The Microsoft Task Scheduler is used to automate the execution of MPE_DL. An administrator level service account was created with permission to run the task even when logged off. Settings selected were the options "Run, whether user is logged on or not" and "Run with highest privileges." This enables continuous loading of MPE data from the incoming REPO files.

2.6. Database challenges

It was found, when loading values from 69,830 HRAP centers listed in each single REPO input file, that the time required to complete one update of the SeriesCatalog per input file was unacceptable. The task of updating the series catalog was taking significantly longer than the actual insertion of data into the DataValues Table of the ODM. This issue was resolved by creating an index on the Series Catalog Table of the ODM. This simple index reduced the Series Catalog Table update time to less than 2 min.

Additional database tuning was needed to expedite data retrieval. While testing a development version of an MPE database, it became clear that tools for retrieving values from the database were failing due to extensive delays. The error was occurring because the Data Values Table accessed by the data download query was too large (over 400 Million values) to respond within a satisfactory time frame. Indexes were again the solution, and this time had to be created on the Data Values Table. Two such indexes were created referencing the Site ID, Variable ID and LocalDateTime fields.

Once applied these two indexes achieved fast response times for retrieval of values from the data tables, but produced one remaining unintended consequence. Rewriting the indexes in order to update them with, for example, the hourly addition of new data, became too time intensive. This led to the current strategy for organizing the entire precipitation dataset. The large archive of historical values is separated from a smaller set of recent values into two individual databases. The larger database of historic values is relatively static and has both of the indexes mentioned above applied to the Data Values Table. Updating the indexes does not create a problem because new values are not frequently added to this set. A smaller database is created to store recent values and receive incoming hourly updates. Because it has a much smaller Data Table, retrieval times are inherently better and do not require the application of the two indexes implemented in the large database. Thus, the precipitation data is separated into two sets. Together, these companion databases are able to store an enormous volume of data, historic up to current, while providing the operational performance needed to support a web service. To measure the performance of the system a sample of

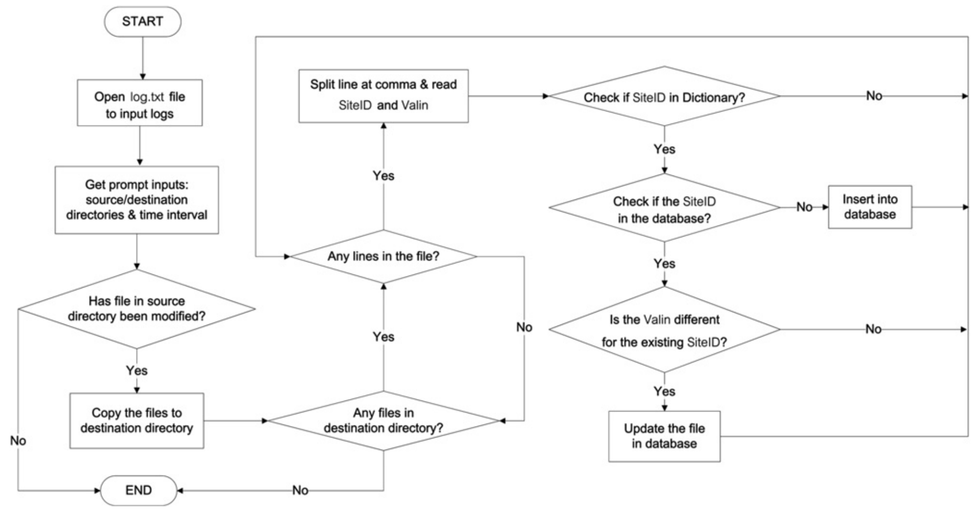

Fig. 3. Flow diagram showing the overall functions of the automated data loading program MPE_DL that was written for data translation and loading.

hourly MPE values was repeatedly downloaded from the large database of historical values for all 151 of the HRAP centers covering Dallas County, TX. The sample size began with one day of data (24 values for each HRAP center, 3624 total values). The test was repeated and the time period of the sample download was incremented each time by one day up to 90 day (326,160 values). The resulting download times began at 2.54 min for a 24 hsample up to 7.11 min for a 90 day sample. The relationship for these results appeared to be linear.

The time period for contents of the recent values database is 12 months. This is observed on the basis of a "water year" rather than a calendar year, meaning values from October 1st to September 30th. Once each year, sometime after October 1st, the content of the recent values database is transferred to the historic values database. Data Table indexes are then rebuilt on the historic values database, and a fresh copy of a recent values database is started.

2.7. Web services

The content of the databases have been published for public access using the CUHASI WaterOneFlow Web Services which is a component of the HydroServer suite.

The WaterOneFlow application uses functions such as GetSites, GetVariables, GetSiteInfo, GetVariableInfo, and GetValues to allow data to be pulled and delivered over the internet in WaterML format. When a request is communicated by a user, WaterOneFlow performs queries of the underlying ODM database, translates the results of each query into the WaterML format, then returns the results.

A separate instance of WaterOneFlow Web Service must be installed and configured for each ODM database created and stored in a host computer. During the configuration process, a unique WSDL is generated for each database and services are registered with the CUAHSI HIS Central Catalog. Registration is largely an automated process during which the HIS Central computer accesses the new WSDL and uses GetSites and GetSiteInfo functions to harvest metadata describing all sites within the new service, as well as details and variables of the data series collected at each site. The final registration step uses a graphical ontology interface called HydroTagger to identify a linkage between the variable ID of the ODM database and the concept ID of the CUAHSI ontology. All contents of this precipitation dataset were linked to the CUAHSI ontology as Hydrosphere/Physical/Flux/Precipitation.

3. Precipitation database content

The web services for hosting MPE data became operational during November, 2010. The original spatial extent of coverage was the WGRFC service, which is the lower magenta colored area shown on Fig. 4, in addition to parts of Mexico. As described above, the best strategy for implementing this dataset as a web service was to build companion datasets: one for historic values and one for recent values. Also note that the MPE values generated by the NWS are issued as hourly and daily products. Thus the outcome was a set of four individual databases, all web accessible. These records of daily and hourly MPE contain the full history of precipitation estimates dating back to 1995 and 1997, respectively. The largest record is the historic hourly set which now contains approximately 8 billion MPE values.

The NEXRAD radars shown in Fig. 1 became operational at different times across the WGRFC area. During the first few years of NEXRAD operation, real MPE values were not generated at all of the 69,830 HRAP grid centers identified in our WGRFC databases. However, in the areas of absent information, the WGRFC REPO files for the earliest few years report zero values rather than a more appropriate blank or null value. This is a very isolated condition, but it leaves the potential for misinterpretation of a zero value as the estimation of no rainfall, while it may actually indicate the absence of an estimated value. As a quality control measure, this project conducted a seek procedure over area and time to identify the presence of such null values and remove the errant zeros from those portions of the database.

Fig. 5 shows the distribution of coverage in Dallas County, Texas as displayed in the CUAHSI client application HydroDesktop. The closely spaced blue icons each represent the center of an HRAP grid cell. Fig. 6 contains Tabular and graphic displays of times series MPE data from Tropical Storm Hermine, Arlington, Texas. Fig. 7 presents the location of MPE currently served for the entire WGRFC service area and the additional regions of Mexico.

The drive space that is required to store the large historic values set of hourly MPE data for the WGRFC is 5.2 Tb. This contains 14 ¼ years of data covering 69,830 sites for a total of approximately 8.7 billion unique values. The difference in storage space between the SQL database volume and the netCDF format is dramatic. Although there is no comparable THREDDS server containing this information in a catalog, it is estimated that the netCDF files to store this information would require only 38.7 Gb.

Fig. 4. Map of service areas for the National Weather Service (NWS).

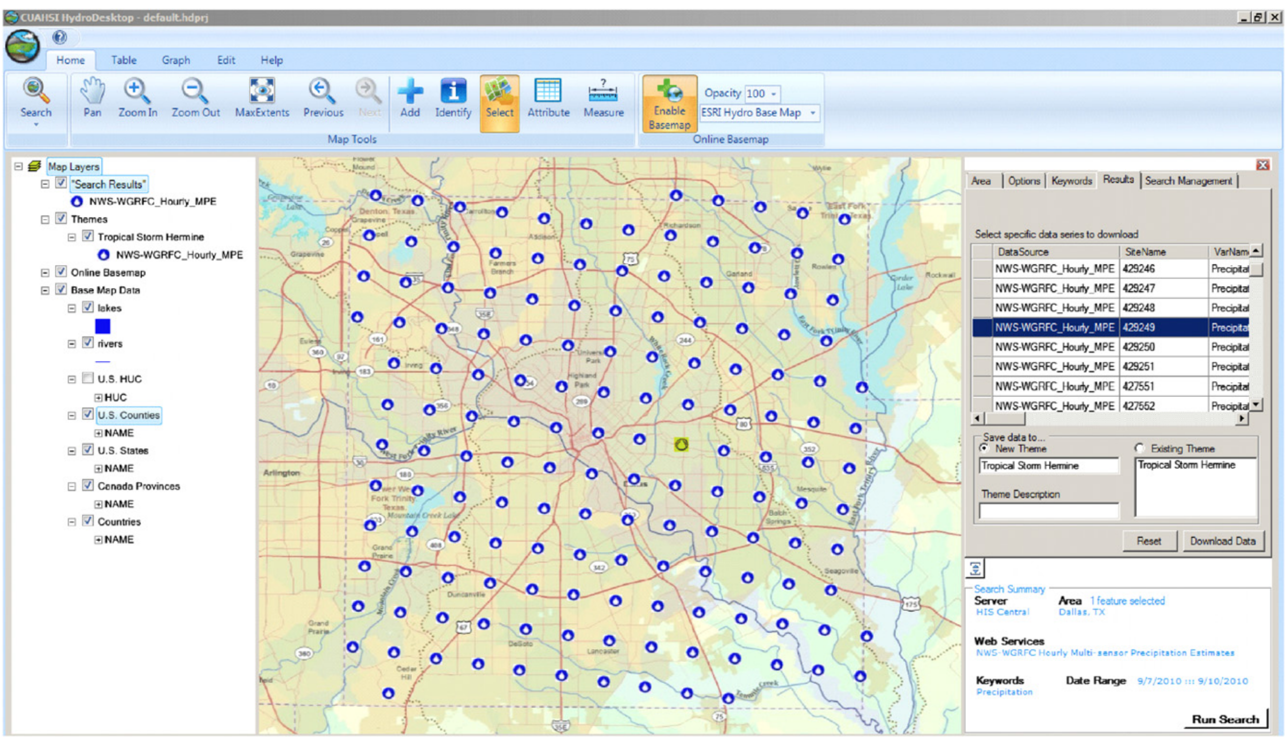

Fig. 5. Distribution of coverage in Dallas County, Texas as displayed in the CUAHSI client application HydroDesktop.

During this past year, the extent of the precipitation dataset has been expanded to include MPE for the neighboring NWS River Forecast Centers: Arkansas-Red Basin (ABRFC) and the Lower Mississippi (LMRFC) (Fig. 1). Additional LDM connections were established with these RFCs to receive the MPE generated at their respective offices for their forecast areas. The services now available for these RFCs provide the recent hourly values, but plans continue to also provide the historic hourly data. Together, the existing services now provide full MPE coverage, updated hourly, of Texas, Oklahoma, Louisiana and Arkansas, along with portions of Mexico, New Mexico, Colorado, Kansas, Missouri, Illinois, Mississippi, Alabama, Kentucky, Tennessee and Georgia. The complete spatial extent of this benchmark dataset is shown in Fig. 8. A summary of the component datasets as well as service addresses are presented in Table 1.

4. Conclusion

The goal of this research has been to provide standardized web service access in time-series format to one of the most valuable hydrologic information resources. The project has established a data platform configured to ingest, store and disseminate MPE values from multiple NWS—River Forecast Centers, currently WGRFC, ABRFC and LMRFC. The data is maintained and archived in a relational database using Microsoft SQL Server. Protocols

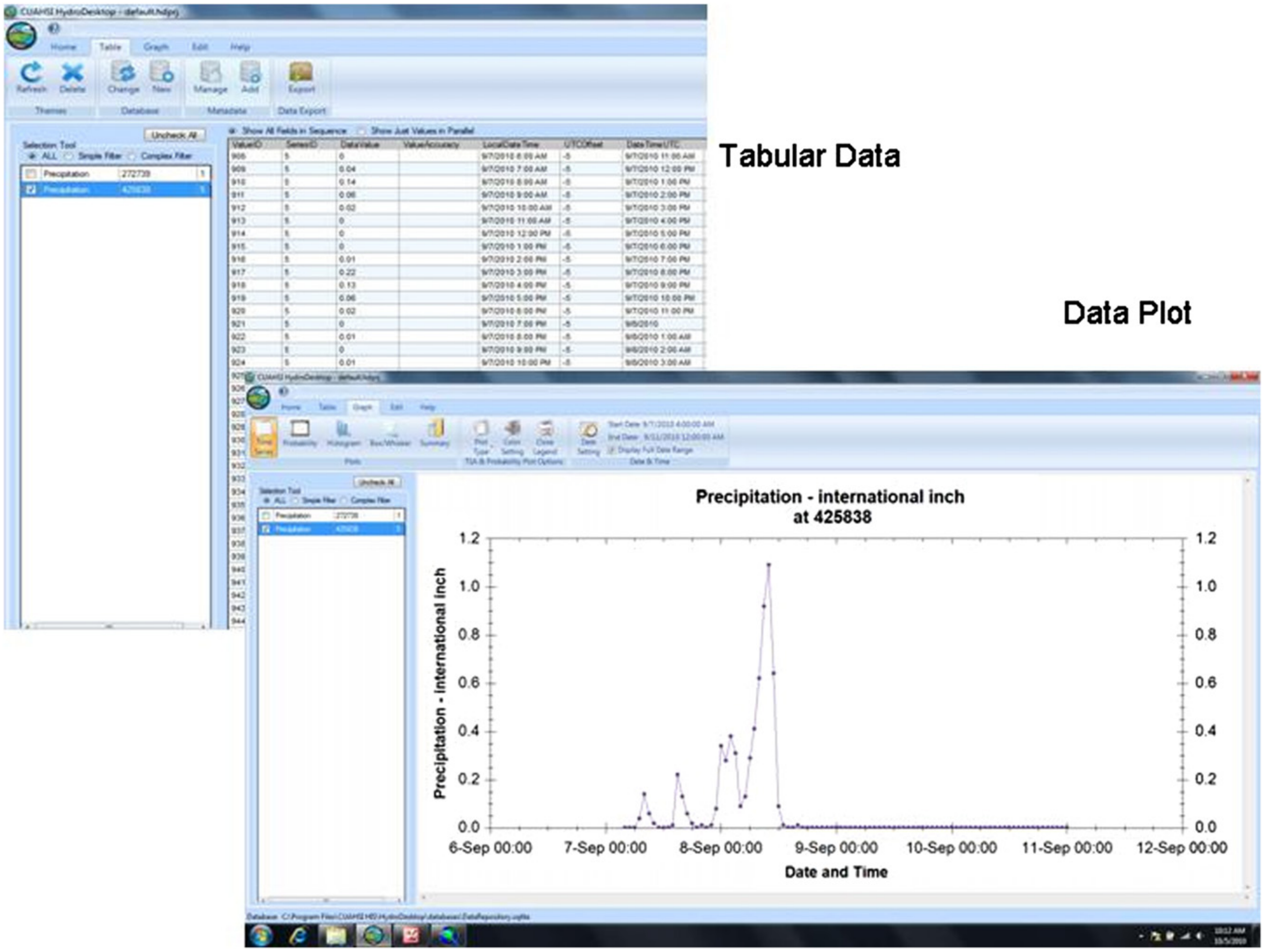

Fig. 6. Tabular and graphic displays of times series MPE data from Tropical Storm Hermine, Arlington, Texas.

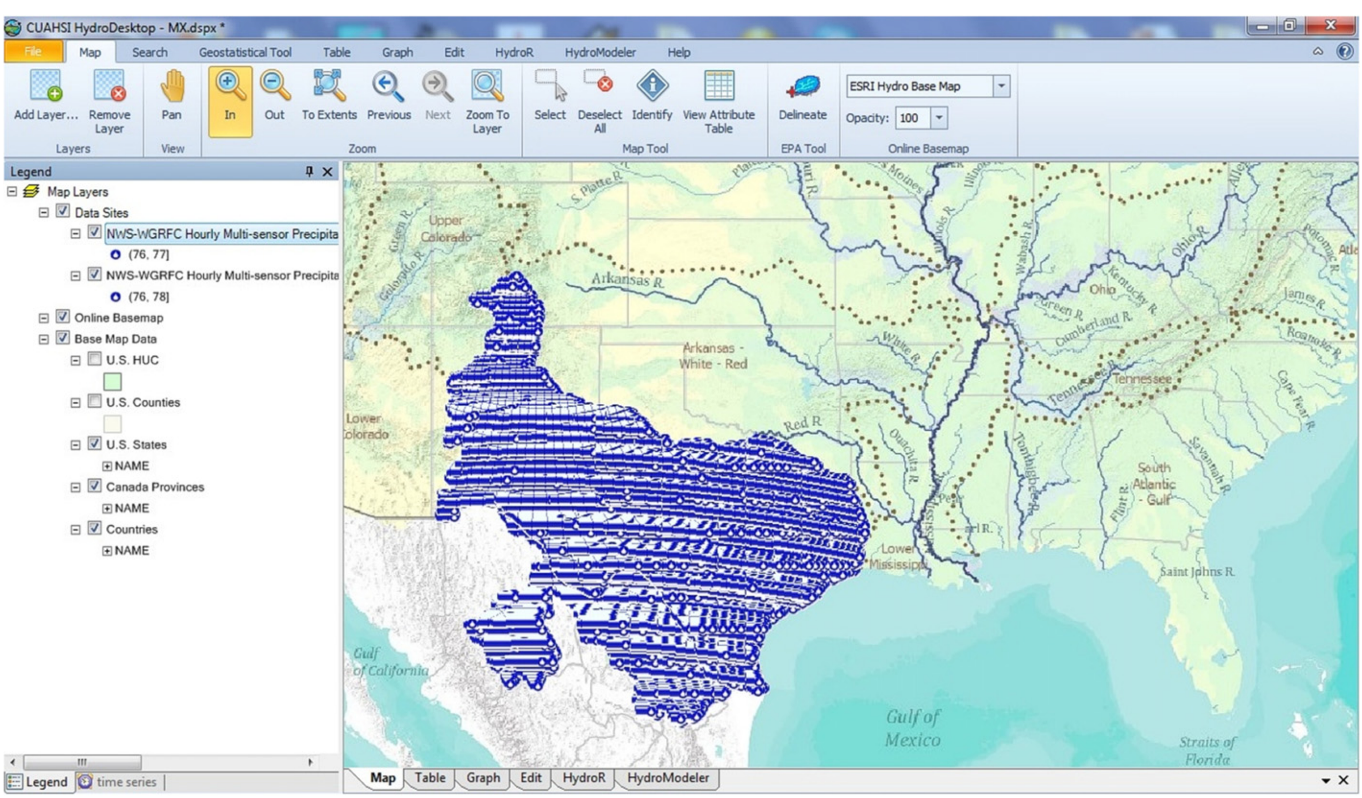

Fig. 7. Coverage of MPE data served for the entire WGRFC service area and the addition regions of Mexico.

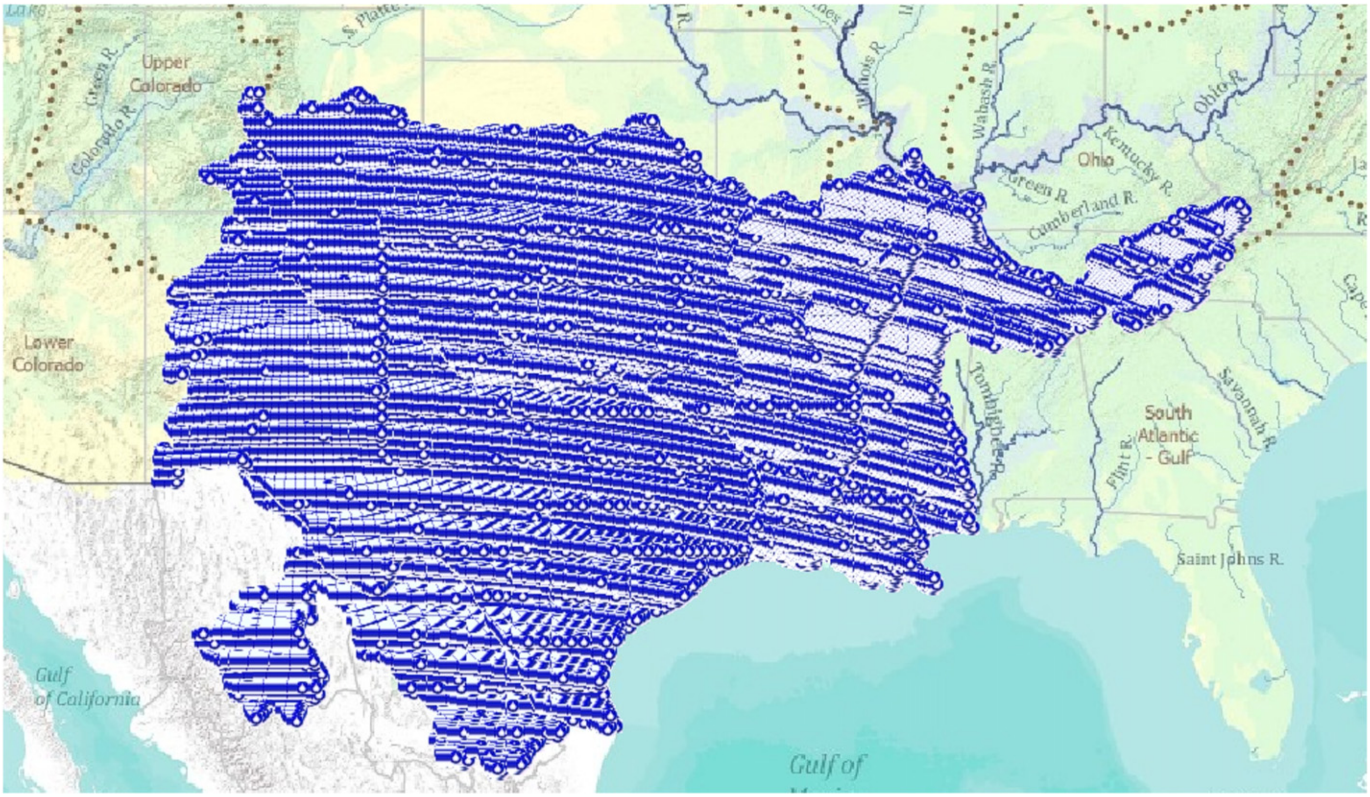

Fig. 8. Spatial extent of hosted hourly MPE values in datasets from WGRFC, ABRFC and LMRFC service areas.

Table 1
Details of the six MPE HIS Services.

Service Title	Network Name	WFS service description	WaterML Service WSDL
Style a			
NWS-WGRFC Hourly Multi-sensor Precipitation Estimates Recent Values	NWS-WGRFC_Hourly_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/190/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_hourly_mpe_recent_values/cuahsi_1_1.asmx?WSDL
NWS-WGRFC Hourly Multi-sensor Precipitation Estimates	NWS-WGRFC_Hourly_MPE	http://hiscentral.cuahsi.org/WFS/187/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_hourly_mpe/cuahsi_1_1.asmx?WSDL
NWS-WGRFC Daily Multi-sensor Precipitation Estimates Recent Values	NWS_WGRFC_Daily_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/189/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_daily_mpe_recent_values/cuahsi_1_1.asmx?WSDL
NWS-WGRFC Daily Multi-sensor Precipitation Estimates	NWS-WGRFC_Daily_MPE	http://hiscentral.cuahsi.org/WFS/188/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_daily_mpe/cuahsi_1_1.asmx?WSDL
NWS-ABRFC Hourly Multi Sensor Precipitation Estimates	NWS-ABRFC_Hourly_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/225/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/ABRFC_Data/cuahsi_1_1.asmx?WSDL
NWS-LMRFC Hourly Multi Sensor Precipitation Estimates	NWS-LMRFC_Hourly_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/226/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/LMRFC_Data/cuahsi_1_1.asmx?WSDL
Style b			
NWS-WGRFC Hourly Multi-sensor Precipitation Estimates Recent Values	NWS-WGRFC_Hourly_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/190/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_hourly_mpe_recent_values/cuahsi_1_1.asmx?WSDL
NWS-WGRFC Hourly Multi-sensor Precipitation Estimates	NWS-WGRFC_Hourly_MPE	http://hiscentral.cuahsi.org/WFS/187/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_hourly_mpe/cuahsi_1_1.asmx?WSDL
NWS-WGRFC Daily Multi-sensor Precipitation Estimates Recent Values	NWS_WGRFC_Daily_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/189/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_daily_mpe_recent_values/cuahsi_1_1.asmx?WSDL
NWS-WGRFC Daily Multi-sensor Precipitation Estimates	NWS-WGRFC_Daily_MPE	http://hiscentral.cuahsi.org/WFS/188/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/nws_wgrfc_daily_mpe/cuahsi_1_1.asmx?WSDL
NWS-LMRFC Hourly Multi Sensor Precipitation Estimates	NWS-ABRFC_Hourly_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/225/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/ABRFC_Data/cuahsi_1_1.asmx?WSDL
NWS-ABRFC Hourly Multi Sensor Precipitation Estimates	NWS-LMRFC_Hourly_MPE_Recent_Values	http://hiscentral.cuahsi.org/WFS/226/cuahsi.wfs?request=getCapabilities	http://hydrodb.uta.edu/LMRFC_Data/cuahsi_1_1.asmx?WSDL

developed by CUAHSI HIS have been employed, including the Observations Database Model (ODM) and WaterML Web Services (WaterOneFlow).

This project has produced the largest structured data set of MPE values, searchable in both time and space, ever compiled. It holds more than 8 billion values beginning in 1995 and up to the most recent hour. Access to the MPE data in a time-series format is important for a wide cross-section of users in the hydrologic community. In essence, the system has created a mesh of "virtual rain gauges" stretched over the landscape for an area of approximately 840,000 square miles shown in Fig. 8. The informational access provided by this benchmark dataset, and the system in which it's sustained, opens possibilities for new developments in hydrologic science and engineering.

Acknowledgments

This work originated from a one-year pilot project supported by the University Corporation for Atmospheric Science (UCAR, Sub-Award #S09-81073).

References

Beran, B., Goodall, J., Valentine, D., Zaslavsky, I., Piasecki, M. 2009. Standardizing access to hydrologic data repositories through web services. In: Proceedings of the International Conference on Advanced Geographic Information Systems and Web Services (GEOWS 2009), IEEE Computer Society, Los Alamitos, CA. Feb. 2009, pp 64–67.

Domenico, B., Caron, J., Davis, E., Kambic, R., Nativi, S., 2002. Thematic Real-time Environmental Distributed Data Services (THREDDS): incorporating interactive analysis tools into NSDL. Journal of Digital Information vol 2 (No 4).

Goodall, J.L., Horsburgh, J.S., Whiteaker, T.L., Maidment, D.R., Zaslavsky, I., 2008. A first approach to web services for the national water information system. Environmental Modelling and Software 23 (4), 404–411, http://dx.doi.org/10.1016/j.envsoft.2007.01.005.

Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., Zaslavsky, I., 2008, A relational model for environmental and water resources data. Water Resources Research 44: W05406, http://dx.doi.org/10.1029/2007WR006392.

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environmental Modelling and Software 24 (8), 879–888, http://dx.doi.org/10.1016/j.envsoft.2009.01.002.

Lawrence, B.A., Shebsovich, M.I., Glaudemans, M.J., Tilles, P.S. 2003. Enhancing Precipitation Estimation Capabilities at National Weather Service Field Offices Using Multi-sensor Precipitation Data Mosaics AMS 83rd Annual Meeting and 19th Conference on IIPS, February 13, 2003, Long Beach, CA. URL: 〈https://ams.confex.com/ams/annual2003/techprogram/paper_54867.htm〉.

Maidment, D.R., 2008a. Bringing water data together. Journal of Water Resources Planning and Management 134 (2), 95–96.

Maidment, D.R., ed., 2008b. CUAHSI Hydrologic Information System: Overview of Version 1.1, Consortium of Universities for the Advancement of Hydrologic Science, Inc, 96 p, 〈http://his.cuahsi.org/documents/HISOverview.pdf〉.

Piasecki, M., Beran, B., 2009. A semantic annotation tool for hydrologic sciences. Earth Science Informatics 2 (3), 157–168, http://dx.doi.org/10.1007/s12145-009-0031-x.

Rew, R.K., Davis, G.P., 1990. NetCDF: An interface for scientific data access. IEEE Computer Graphics and Applications, 76–82.

Sheth, A.P., Larson, J.A., 1990. Federated database-systems for managing distributed, heterogeneous, and autonomous databases. Computing Surveys 22 (3), 183–236.

Skonnard, A. 2003. Understanding WSDL. from 〈http://msdn.microsoft.com/en-us/library/ms996486.aspx〉.

Tarboton, D.G., Jeffery S. Horsburgh, J.S., Maidment, D.R. 2008. CUAHSI Community Observations Data Model (ODM), Version 1.1, Design Specifications.

Thompson, M. 2010. [Hydrologist, National Weather Service—West Gulf River Forecast Center] Personal Communication.

Unidata. 2011. LDM Factsheet. From 〈http://www.unidata.ucar.edu/software/ldm/ldm-current/factsheet.html〉.

Whitenack, T. 2010. CUAHSI HIS Central 1.2 Web based Data Service Repository for HIS Web Services: Consortium of Universities for the Advancement of Hydrologic Science, Inc.

Zaslavsky, I., Valentine, D., Whiteaker, T. 2007. CUAHSI WaterML OGC Discussion Paper OGC 07-041r1. Version 0.3.0.