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

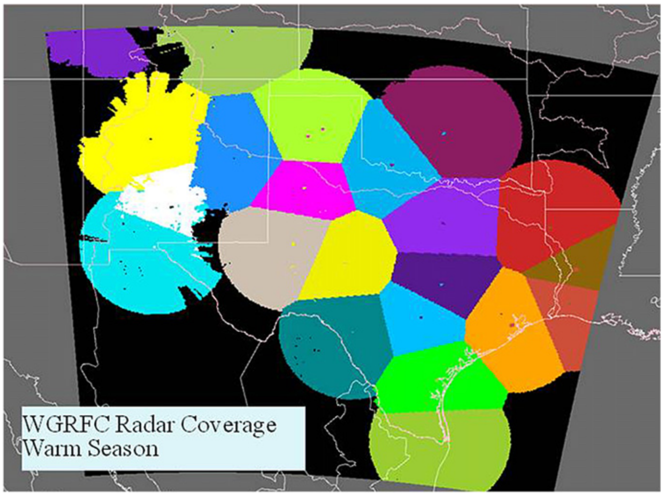

Fig. 1. Mosaic of NEXRAD radars utilized by the NWS WGRFC (NWS).

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