Southwest Watershed Research Center Data Access Project

M. H. Nichols[1] and E. Anson[1,2]

Received 27 October 2006; revised 6 June 2007; accepted 11 October 2007; published 26 February 2008.

[1] Hydrologic data, including rainfall and runoff data, have been collected on experimental watersheds operated by the U.S. Department of Agriculture Agricultural Research Service (USDA-ARS) in southern Arizona since the 1950s. These data are of national and international importance and make up one of the most comprehensive semiarid watershed data sets in the world. The USDA-ARS Southwest Watershed Research Center has recently developed an electronic data processing system that includes an online interface (http://tucson.ars.ag.gov/dap) to provide public access to the data. The goal of the system is to promote analyses and interpretations of historic and current data by improving data access. Data are collected from sensors in the field and are transmitted to computers in the office. The data are then processed, quality checked, and made available to users via the Internet. The publicly accessible part of the system consists of an interactive Web site, which provides an interface to the data, and a relational database, which is used to process, store, and manage data. The system was released to the public in October 2003, and since that time the online data access Web site has received more than 4500 visitors.

Citation: Nichols, M. H., and E. Anson (2008), Southwest Watershed Research Center Data Access Project, *Water Resour. Res.*, *44*, W05S03, doi:10.1029/2006WR005665.

1. Introduction

[2] Since its establishment in 1953, the Southwest Watershed Research Center (SWRC) in Tucson, Arizona, has collected, processed, managed, and disseminated high-resolution, spatially distributed hydrologic data in support of the center's mission. Data management at the SWRC has evolved through time in response to new computing, storage, and data access technologies. In 1996, the SWRC initiated a multiyear project to upgrade rainfall and runoff sensors and convert analog systems to digital electronic systems supported by data loggers. This conversion was coupled with radio telemetry to remotely transmit recorded data to a central computer, thus greatly reducing operational overhead by reducing labor, maintenance, and data processing time. A concurrent effort was initiated to improve access to SWRC data by creating a system based on a relational database supporting access to the data via the Internet. An SWRC team made up of scientists, IT specialists, programmers, hydrologic technicians, and instrumentation specialists was formed. This effort is termed the Southwest Watershed Research Center Data Access Project (DAP).

[3] The goal of the SWRC DAP is to efficiently disseminate data to researchers; land owners, users, and managers; and to the public. Primary access to the data is provided through a Web-based user interface. In addition, data can be accessed directly from within the SWRC network. The first priority for the DAP was to assimilate and make available rainfall and runoff data collected from two instrumented field sites, the Walnut Gulch Experimental Watershed (WGEW) near Tombstone, Arizona, and the Santa Rita Experimental Range (SRER) south of Tucson, Arizona

[4] The objectives of this paper are to briefly describe instrumentation and to provide an overview of the SWRC DAP emphasizing the data management system. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

2. Brief Instrumentation Description

[5] Hydrologic data have been collected at high spatial and temporal resolutions on the WGEW since 1953, and at 8 ARS watersheds within the SRER since 1975. Currently, data are collected at 125 instrumented sites on the WGEW (P. Heilman et al., GIS database, Walnut Gulch Experimental Watershed, Arizona, USA, submitted to *Water Resources Res.*, 2006) and 18 instrumented sites on the SRER. Precipitation is measured at 88 sites throughout the WGEW watershed, with a density of 1.7 gauges km^{-2} (0.7 $mile^{-2}$) [*Goodrich et al.*, 2008]. There are eight rain gauges at the SRER. Within each gauge, water collected during a precipitation event is weighed by a platform mounted container using a precision, temperature compensated load cell. Voltage output from the load cell changes in response to water accumulating in the container. A data logger is programmed to sample the voltage every second and record the average at 1 min intervals. To minimize data storage requirements and transmission time, only times and voltages commensurate with precipitation detectable to 0.25 mm (0.01 in) are recorded. The maximum capacity of the rain gauge is 200 mm (8 in). The electronics, data

[1]Southwest Watershed Research Center, Agricultural Research Service, U.S. Department of Agriculture, Tucson, Arizona, USA.
[2]Now at Zardick Computing, Tucson, Arizona, USA.

0043-1397/08/2006WR005665$09.00

Table 1. General Data Type Descriptions

Data Types	Description	Current Status
Quality checked	Includes SWRC metadata and quality reviewed data from SWRC core monitoring activities: data from 1953 through the current year (precipitation, runoff, sediment, and meteorological) online publications, citations, and abstracts	Precipitation and runoff from WGEW and SRER available through 2007 via standard and graphical interfaces
Custom	Includes preliminary daily precipitation, runoff, and stock tank reports available at WGEW field office, and monthly printed summaries sent via U.S. mail	Available by request through WGEW field office
Preliminary research data	Includes project specific data from research directed by SWRC scientist or students monitoring activities and associated metadata supported by SWRC funds, and resources contributed by SWRC-affiliated research programs	Contact individual project scientist
Sensitive data	Includes data sets subject to proprietary or copyright restrictions; data requests require prior approval by the SWRC research leader	Currently none

logger, and radio/modem components are housed in a metal belowground cylinder, thus reducing vandalism, lightning interference, and temperature effects.

[6] Runoff is monitored at 11 large flumes sites and 10 small flume sites within the WGEW. There are 8 small flumes at the SRER. In addition, there are 10 stock tanks within the WGEW instrumented with stilling wells containing floats to measure water-surface depth. The analog system to measure water depth was converted to a digital system by attaching a precision linear potentiometer to the output gear shaft of the water-level recorders. The voltage output from the potentiometer is collected by a data logger which averages 1-s samples at 1-min intervals and records flow data (time and voltage) only when a minimum depth threshold has been exceeded (0.003 m (0.12 in) at small flumes, 0.015 m (0.59 in) at large flumes and stock tanks).

[7] Two intensive study sites within the WGEW, one characterized by grass cover (containing two), and one characterized by brush cover (containing 6 subwatersheds) are monitored with meteorological, soil moisture and temperature and energy flux measurements. These data sets will be incorporated in the DAP in the future.

[8] Shortly after midnight on a daily basis, data from the WGEW sites are downloaded automatically via radio and are transmitted to a computer at the SWRC Tombstone field office. Daily radio transmission time and file size vary with weather conditions on the watershed. In the absence of significant rainfall or runoff, radio contact and data transmission can be accomplished in approximately 1.5 h. During the summer "monsoon season" these tasks can take over 4 h. Data are archived and copied over the Internet via a secure virtual private network to an SWRC server residing in Tucson. Data from the SRER sites are transmitted daily via cellular modem to the SWRC office in Tucson.

3. Data Management System

[9] The SWRC data management system provides a framework for storing, organizing, archiving, and retrieving data associated with SWRC research. The database and Web site were designed by SWRC staff, who also wrote the data processing code. The goals of the data system are (1) to promote analyses and interpretations of data by improving access and (2) to meet our obligation to make data available to the public.

3.1. Data Users

[10] A critical first step in the overall design of the DAP was to identify potential users and their data needs, including timeliness of data delivery and data quality (Table 1). The primary data users are expected to be researchers and scientists needing access to long-term, quality-checked data sets. Data in this category include precipitation and runoff measurements that have been processed using standard SWRC data reduction procedures, and are made available via the Web site with updates posted approximately within a week after hydrologic events. A second user group is made up of ranchers and local landowners who require information such as rainfall amount and the depth of runoff into stock tanks for making management decisions. These users can access preliminary data within 24 h of recording, and they can also request a printed report summarizing preliminary information. These data have not been reviewed for quality assurance. A third category of data users is made up of scientists and students collecting research specific data that will be interpreted and published. These data are expected to be available following publication of research results, and data quality and verification are the responsibility of the principal investigator on the project. The collection of sensitive data is expected to be a rare occurrence, however, access to such data will be handled on a case by case basis.

3.2. General Layout and Implementation

[11] A significant effort went into designing the overall data table structure and logic, including table definitions and relationships. The overall goal was to create a simple read only online analytical processing (OLAP) system with a structure that would allow for synthesis of information. A key challenge to database design was accommodating the historic analog data (which existed in text files and on paper), the current analog data (which is being run in parallel with the new sensor network to verify the data sets), and the electronic data stream generated by the new sensor network. Rainfall and runoff data historically were collected through analog pen traces on charts driven by mechanical clocks [*Brakensiek et al.*, 1979]. Charts were digitized and data were entered into text files. These rainfall and runoff data have been stored and accessed in an evolving system of computer hardware and software.

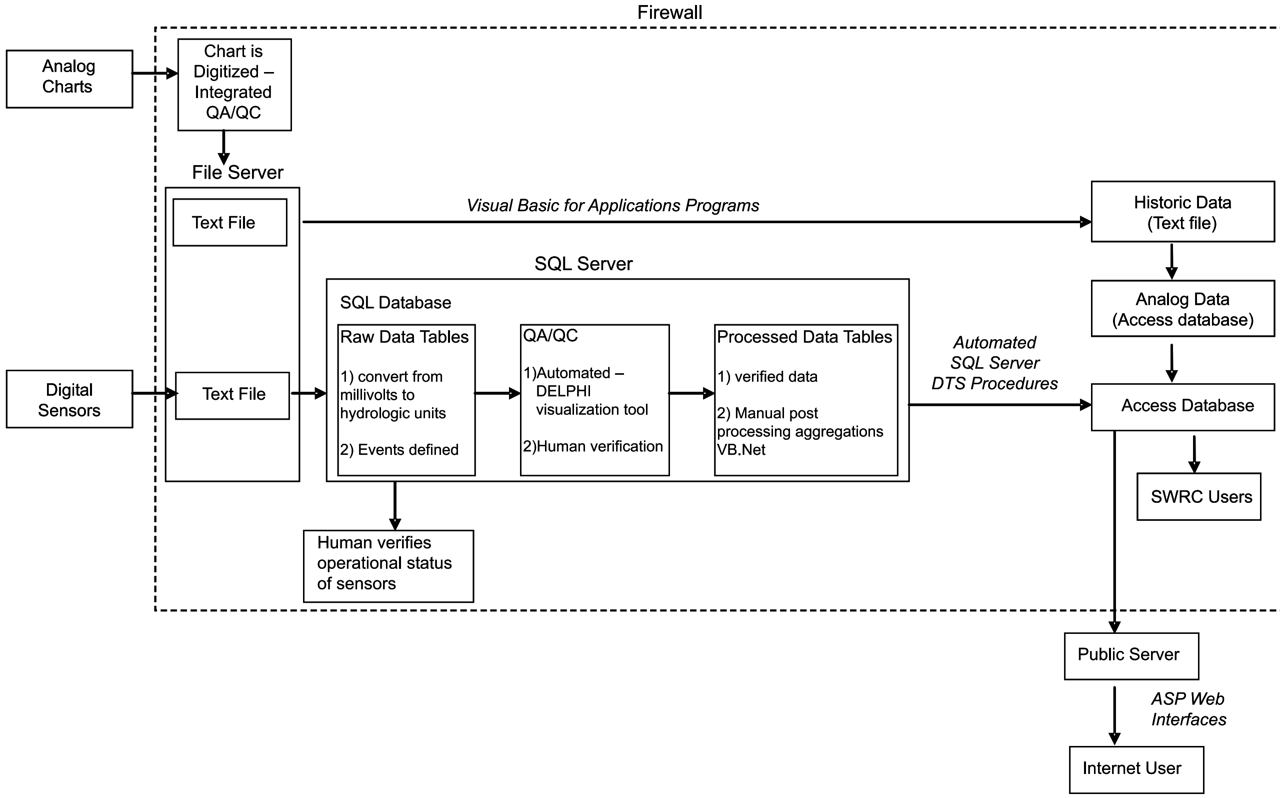

Figure 1. General data flow and implementation diagram.

[12] The general data flow is shown graphically in Figure 1. Data collected by both the digital sensors and analog charts arrives at a file server behind the SWRC firewall, where subsequent processing follows two distinct paths. The text file data from the electronic sensors are parsed into a collection of "raw data" database tables. These data are processed into a collection of "processed" database tables in a Microsoft SQL Server 2000 database. The text file data from the analog charts are added to a Microsoft Access database which closely matches the text file structure. Finally, postprocessing algorithms are implemented to populate an Access database which contains both historic and digital data. The file-based nature of MS Access databases facilitates easy data distribution using a variety of methods. Processed data in the Access database are made available to the public via the SWRC public Web server, by direct database connection within the SWRC network, or by request on a CD. The public processed database is updated approximately weekly.

[13] An important feature of data management is access to information describing the status of instruments and data transfers. This information is used by field personnel to maintain the instrumentation network. Sensor, battery, data logger, and data transfer status are recorded within the DAP SQL database on a daily basis in a set of tables created specifically to facilitate maintenance of the instrumentation network. Problems such as mechanical instrument malfunctions, sensor and data logger malfunctions and data transfer gaps can be identified. In addition, disturbances by animals, wind, and vandals can be identified. SWRC field operations personnel review battery voltages, and other data pertaining to instrument status through a Web-based interface accessible within the SWRC network.

3.3. Server Hardware and Software

[14] Initial data collection is accomplished through Campbell Scientific software running on MS Windows XP PCs in the Tombstone field office and in Tucson. Data in raw (millivolt) format are written to text files that are organized by date and instrument using the file system directory structure. After all the data for the day have been collected, an MS DOS batch file copies the files to a RAID 5 disk drive array on a file server in Tucson running MS Windows Server 2000. A program on the file server runs hourly to check if the data transfer is complete. When the program detects a completed data transfer it parses the text files, processing the data and populating tables of the SWRC database. The SWRC database resides on a DELL File Server running MS Windows Server 2000 and MS SQL Server 2000. Processing these data includes converting millivolt values to hydrologic units of depth and grouping rainfall and runoff data into "events." Both rainfall and runoff events are defined at the instrument level, such that data from other instruments is not considered in determining the occurrence of an event. The start of any precipitation or runoff data marks the beginning of an event. An event ends when no new precipitation or runoff data is recorded for a period of an hour or longer. Events created by this data processing program are marked as either "unchecked" or "open." An open event is one that has not ended by the time the daily data was collected from the data logger. The remaining events are marked as unchecked indicating they represent a complete event and are ready for further QA/QC

checking and processing. SWRC personnel developed the software to parse and process data using MS Visual Basic .NET 2002.

[15] The precipitation and runoff events are then quality checked by visual inspection. A visualization tool that runs on Windows XP desktop computers was developed using Borland Delphi. This program queries the database for "unchecked" events and displays them graphically for the user. In addition to displaying a graph of the time series of an event for a particular instrument, it also displays a color coded map which represents daily summary precipitation and runoff over the watershed. If the time series graph looks typical and the magnitude and duration of the event are judged to be within the range of expected values on the basis of the daily summaries across the watershed, the event is marked as "verified." Otherwise the event is marked as "not good." The program also provides methods for correcting common problems. For example, measuring runoff using the flumes requires that a direct hydraulic connection between the flume intake plates and a vertical stilling well where water level is measured is maintained. The sediment laden flows that characterize runoff on the WGEW typically deposit sediment through the intake plates thus affecting the hydraulic connection resulting in the loss of data during the later part of the flow recession. A method for fitting a curve to estimate the recession when these data are lost was developed in house on the basis of using a French curve to extend the measured hydrograph. Base flow is not common in the SWRC data, with one exception. Base flow was common at flume 2 on the WGEW through 1978, and because the program can be used to distinguish between individual runoff events, it can also be applied to separate base flow from flood flows during data processing. The original data associated with all modified runoff events are stored in the database and can be recovered if necessary. The program can also be used to reopen events previously marked as "verified" or "not good" to reevaluate or modify their status.

[16] Verified data are periodically aggregated to produce summarized hydrologic data, such as daily rainfall totals. The aggregated data are stored in data tables on the database server. A postprocessing program is manually run on the database server computer to update the aggregate information and incorporate events whose status (unchecked, verified, or not good) has changed. This program was also written using Visual Basic .NET. About once a week, MS SQL Server 2000 procedures are run to export the event and aggregate data to an Access database that contains data collected with analog instruments.

[17] To make the data accessible on the Internet this Access database is manually copied to a Web server. The Web server is a Dell Server machine running MS Windows Server 2000 operating system that is located outside of the internal SWRC network. The interactive Web interface for the DAP was developed using Visual Basic Script and MS Active Server Pages (ASP). The ASP scripts use MS Active Data Objects to query the Access database and return data to the end user.

3.4. Extracting and Viewing Data

[18] Access to SWRC and SRER data can be found at http://www.tucson.ars.ag.gov/dap. Both the historic analog and current digital data are available. Access to the data is provided by two methods: (1) a standard forms-based method and (2) a geographic information systems (GIS)-based graphical method.

[19] The standard method is implemented using HTML forms and ASP. The ASP interface displays HTML forms that allow the Internet user to select date ranges, instruments, aggregation, units, summary statistics and output format. When the form is submitted, ASP runs Visual Basic Script that queries the Access database and then processes and formats the data that is subsequently displayed for the user in their chosen format. This interface accommodates a wide range of data selection and retrieval choices and can provide the requested data in HTML, MS Excel and text formats.

[20] The standard forms-based method of requesting data can be efficiently used by those familiar with the field sites and the instrumentation network. In 2005, graphical access to the data was developed using ESRI ArcIMS. The graphical interface integrates high-resolution orthophotos, Landsat satellite images, and GIS data layers containing watershed boundaries, channel networks, and point locations of flumes and rain gauges. After the user selects instrumented sites through the graphical display, the site parameters are used to populate the standard form, and data are displayed in response to submission.

3.5. Additional Administrative Features

[21] The SWRC has developed a voluntary data use agreement. Web site usage is monitored by keeping a count of the number of visits to the Web site, but no additional user data are collected. Data security is maintained through protocols established in compliance with the USDA Office of the Chief Information Officer (OCIO) protocols. The SWRC computer network resides behind a firewall, and access to internal servers is limited to on-network computers.

[22] The ARS experimental watersheds are a valuable source of high-quality, spatially distributed data collected over long time periods. These data are a critical resource, and the SWRC has a strong interest in promoting their use in collaborative interdisciplinary research. All data available through the SWRC DAP Web site are in the public domain, and are not restricted by copyright. Funding for these data sets was provided by the United States Department of Agriculture, Agricultural Research Service (USDA-ARS). The SWRC requests that authors forward a copy of manuscripts based on the data, and acknowledge USDA-ARS SWRC.

4. Conclusion

[23] Data collected at the WGEW and SRER are a valuable resource to researchers; land owners, users, and managers; and to the public. The SWRC DAP project has made access to these data easy. The DAP is, and will continue to be, a work in progress. The foundation is established to add additional data sets including, meteorologic, soil moisture, GIS, topographic survey, vegetation, remote sensing and other data. We anticipate that computer systems and technologies will evolve, and the DAP will evolve in response. However, the primary goal of the project to make data available has been met, and expanded access to the data will be a significant contribution to natural resources research throughout the world.

[24] **Acknowledgments.** Since 1953, numerous individuals have contributed to instrumentation, data collection, data processing, data integrity, and data access. Chris Nichols, Saricorp.com, provided substantial assistance with database design and project integration. Their contributions, and the continued contributions of Jason Wong and Carl Unkrich, are gratefully acknowledged.

References

Brakensiek, D. L., H. B. Osborn, and W. J. Rawls (1979), Field manual for research in agricultural hydrology, *USDA Handb. 224*, 550 pp., U.S. Dep. of Agric., Washington, D. C.

Goodrich, D. C., T. O. Keefer, C. L. Unkrich, M. H. Nichols, H. B. Osborn, J. J. Stone, and J. R. Smith (2008), Long-term precipitation database, WGEW, Arizona, USA, *Water Resour. Res.*, doi:10.1029/2006WR005782, in press.

Stone, J. J., M. H. Nichols, D. C. Goodrich, and J. Buono (2006), Long-term runoff database, Walnut Gulch Experimental Watershed, Arizona, USA, *Water Resour. Res.*, doi:10.1029/2006WR005733, in press.

E. Anson, Zardick Computing, 3445 S. Crown Dancer Drive, Tucson, AZ 85730, USA.

M. H. Nichols, Southwest Watershed Research Center, ARS, USDA, Tucson, AZ 85719, USA. (mary.nichols@ars.usda.gov)