STEWARDS Watershed Data System: System Design and Implementation

J. L. Steiner, E. J. Sadler, G. Wilson, J. L. Hatfield, D. James,

B. Vandenberg, J.-S. Chen, T. Oster, J. D. Ross, K. Cole

Abstract. *A web-based, data retrieval application was developed (Sustaining the Earth's Watersheds, Agricultural Research Data System, or STEWARDS) as part of the USDA-ARS Conservation Effects Assessment Project (CEAP) to increase the availability and accessibility of scientific data to the research community. The STEWARDS application is GIS-based and couples temporal and spatial aspects of data collected from each site within a watershed. The STEWARDS database and software design accommodates research data with heterogeneous characteristics and format, and captures rich descriptive information that is important to understand the data from complex, dynamic research programs. The database includes soil, water, climate, land management, and socio-economic data from multiple watersheds across the U.S. and can provide data commonly needed for hydrologic modeling and assessments. The release of STEWARDS marks an advance in the research capacity for the ARS research watershed network by improving access to well-documented and consistently organized data and is becoming the prototype of additional data and software designs for other ARS research projects.*

Keywords. *CEAP, Conservation Effects Assessment Project, Databases, Data system design, Internet-based data system, User interface.*

An increasing number of web-based data systems are becoming available in natural resources research fields such as geosciences (geoinformatics), hydrology (hydroinformatics), ecology (ecoinformatics), and biology (bioinformatics). Application of software and database technologies can overcome problems of fragmentation, inadequate documentation, and cumbersome manipulation of agroecological research data (van Evert et al., 1999a). Research teams have demonstrated the practicality and benefits of implementing a database at the onset of complex, multi-site, multi-year agroecological experiments (van Evert et al., 1999b; Scott and Lord, 2003). However, there is also a need to provide better data management alternatives for historical and ongoing research and monitoring sites.

Watershed studies require a multitude of spatial and temporal data that may be highly diverse across watersheds. Often, data sets have been collected by different organizations within the same geographic area, and the value of the data is enhanced if they can be compiled for access at a common site. Patino-Gomez et al. (2007) described development of a database for the Rio Grande/Bravo river basin, incorporating data from different sources in the U.S. and Mexico to support analysis and resolution of water issues. Carleton et al. (2005a, 2005b) developed a relational database to improve the efficiency of managing data for watershed monitoring and assessment projects that includes quality control, application of conversion factors, retrieval of data for applications, and data comparisons among multiple sites. Lane (1997) described the Environmental Change Network data system, which provides a variety of environmental data from 12 terrestrial and 45 freshwater sites in the U.K.

The USDA Agricultural Research Service (ARS) has built a web-based geospatial database application to address data management for research efforts covering decades of measurements, multiple research initiatives, and multiple locations. This database application, called Sustaining the Earth's Watersheds, Agricultural Research Data System (STEWARDS), supports the USDA-ARS Conservation Effects Assessment Project (CEAP) by bringing decades of heterogeneous data into an organized, well-documented database (ARS, 2008). The database accommodates measurement data for weather, soils, hydrology, water quality, land use, land management, and socio-economics, as well as survey data, spatial GIS layers, metadata, and descriptive text for specific watersheds, and was developed to provide improved accessibility and utility of the data, as well as to enhance visibility of this network of watersheds to a broader

Submitted for review in July 2008 as manuscript number SW 7570; approved for publication by the Soil & Water of ASABE in August 2009.

References to products are for information purposes only and do not constitute an endorsement by the USDA or the authors.

The authors are **Jean L. Steiner,** Supervisory Soil Scientist, USDA-ARS, El Reno, Oklahoma; **E. John Sadler, ASABE Member,** Supervisory Soil Scientist, USDA-ARS, Columbia, Missouri; **Greg Wilson,** Information Technology Specialist, USDA-ARS, Beltsville, Maryland; **Jerry L. Hatfield,** Supervisory Plant Physiologist, USDA-ARS, Ames, Iowa; **David James,** Geographic Information Specialist, USDA-ARS, Ames, Iowa; **Bruce Vandenberg,** Information Technology Specialist, USDA-ARS, Fort Collins, Colorado; **Jin-Song Chen,** Former Research Associate, USDA-ARS, El Reno, Oklahoma; **Teri Oster,** Information Technology Specialist, USDA-ARS, Columbia, Missouri; **John D. Ross,** Information Technology Specialist, USDA-ARS, El Reno, Oklahoma; and **Kevin Cole,** Physical Scientist, USDA-ARS, Ames, Iowa. **Corresponding author:** Jean L. Steiner, USDA-ARS, 7207 West Cheyenne Street, El Reno, OK 73036; phone: 405-262-5291; fax: 405-262 0450; e-mail: jean.steiner@ars.usda.gov.

audience in the natural resources research and management communities.

Steiner et al. (2008) discussed the challenges of designing a system for heterogeneous data and the project management during design and development of STEWARDS, and Sadler et al. (2008) provided an overview from the perspective of users retrieving data, researchers providing data to STEWARDS, system administrators managing the system, and software developers maintaining the application's code. The objective of this article is to describe the functionality and content of STEWARDS, the approaches used to handle heterogeneous data, and the operational features. Additionally, we will highlight unique features and approaches that might guide other watershed database development efforts.

Functionality and Content

A core objective in developing STEWARDS was to provide increased accessibility, utility, and visibility to a national watershed research program that supports research in numerous watersheds. The approach taken was to provide a one-stop site to gather information and data from various watersheds. A conceptual schematic of the STEWARDS architecture (fig. 1) shows a web-based client-server system with four key components:

- A centralized system with servers and controls, applications, and management and science teams. The server controls provide the web applications, services, and rules of system task flow. The infrastructure includes relational database management servers, a web server, an application server, and data/map servers.
- User interfaces: The server takes a web user's request to an application program (e.g., data search, access, visualization, or download) and sends the resulting pages back to the user at remote sites.
- Data sources: Researchers and data managers at watershed sites are data providers and are responsible for preparing data for upload.
- Data storage: Relational databases that include descriptive information, measurement data, images, GIS layers, and metadata.

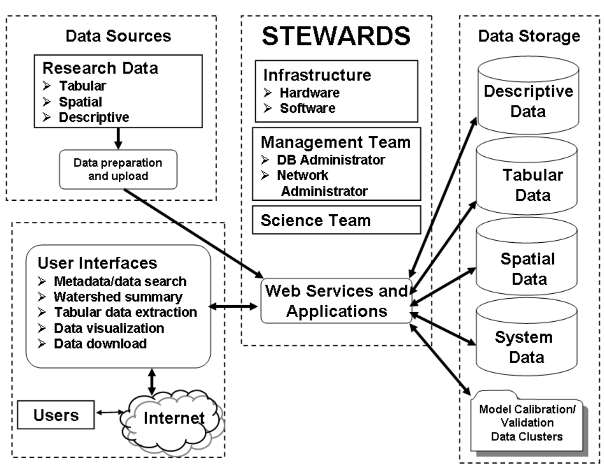


Figure 1. Conceptual representation of STEWARDS with key components bounded by dashed lines.

Handling Heterogeneous Data

A key challenge was to provide adequate flexibility to accommodate diverse data from the watersheds, while presenting the data consistently to users. Each watershed provides some combination of measured data (e.g., weather, soils, hydrology, water quality); land use, management, and socioeconomic data; and GIS layers, but the data structures vary across different watersheds and often within a particular watershed over time. Measurement data may be represented by single observations (e.g., soil properties), time series at a particular location (e.g., weather, hydrology, water quality), or spatial measurements presented in GIS layers (e.g., land use). The GIS database contains the watershed boundary, location of measurement points, and spatial features such as a digital elevation model, stream network, soil map, land use, or land management data.

Specific types of data available in the STEWARDS database include:

- Descriptive data (i.e., watershed descriptions, site descriptions).
- Tabular measurement data.
- Spatial measurement or characterization data with associated metadata.
- Collections of files for model calibration/validation analyses (i.e., input, output, and executable files for a published analysis available for download as a .zip directory).

Table 1 summarizes the data sets provided for two research subwatersheds in the Upper Washita River basin, in Oklahoma, that are available in STEWARDS or will be made available in the future. These two watersheds provide contrasts of short and long periods of record representative of watersheds providing data to STEWARDS. The Little Washita River watershed has been the focus of ARS research since 1961, addressing a multitude of research questions. There is a wealth of legacy data, for which details about methodology may be sparse. Methodology and monitoring sites have changed through the years with differing research objectives, personnel, resource availability, and available technologies. Research in the Fort Cobb Reservoir was initiated by ARS in 2004, as part of the CEAP, and therefore much of the data are only now being prepared for initial reporting, with subsequent data delivery to STEWARDS. Methods for the Fort Cobb watershed research are easier to prepare because they can be described in detail by current researchers, and the methods have been stable over the period of record, to date.

Rich Descriptive Data

Watershed and Site Descriptions

For each watershed, the watershed team develops a watershed description that can include a variety of text, tables, figures, or photographs to describe watershed characteristics, issues of concern, research focus, and partnerships in the watershed.

The watershed teams also develop information to describe measurement sites. The database includes several tables: SITE contains the main site description, SITE_INSTRUMENTS contains instrument names and descriptions with one-to-many cardinality with analytes (i.e., one instrument may be used to measure many analytes), and SITE_PICTURES contains site images in a binary format. When a user selects a measurement site in the watershed, STEWARDS generates a site description using Crystal Re-

Table 1. Data catalog for watersheds in the Upper Washita River basin of Oklahoma.

Name	Type of Data	Content	Status[a]
Upper Washita River watershed (HUC[b]: 11130302)	Descriptive	Watershed description	S
The Upper Washita River has mixed agricultural land use, including row cropping, small grains, prairie rangelands, and pastures. Red cedar invasion into native rangelands is a major concern. Sedimentation and nutrient loading from agriculture are major concerns, as well as channel instability in some of the tributaries. The infrastructure of several rural communities (bridges, a water treatment plant) is threatened, and water bodies are impaired for municipal water supply, recreation, and fish and wildlife.			
Little Washita River subwatershed (HUC: 11130302050)	Descriptive	Watershed description	S
		Field and laboratory methods for observation data	S
The ARS and partners have conducted research in this watershed from 1961 to the present. The watershed's topography, ecosystems, and agriculture are typical for the southern Great Plains. The research focus is validation and improvement of hydrologic models, particularly the Soil and Water Assessment Tool (SWAT), and evaluation of environmental effects of agricultural conservation and management.	Spatial - characterization layer with appropriate metadata	Watershed boundary	S
		Digital elevation model	S
		Stream network	S
		Soils	S
		Geology	S
		Land use 1994	L
		Land use 2005	Q
		Channel stability (Simon and Klimetz, 2008)	Q
		Impoundments	Q
	Spatial - measurement site layer with appropriate metadata	Weather stations	S
		Stream gauges	S
		Soil sampling sites	S
		Monitoring wells	L
		Characterization wells	L
	Observation - time series	Weather	S
		Stream flow	S
		Groundwater depth	L
	Observation - event based	Suspended sediment at gauging stations	S
	Observation - ad hoc	Soil texture and density	S
Fort Cobb Reservoir subwatershed (HUC: 1130302160, above the dam)	Descriptive	Watershed description	Q
		Field and laboratory methods for observation data	Q
Research has been conducted in the Ft. Cobb Reservoir watershed from 2004 to the present. The Oklahoma Conservation Commission supported a Water Quality Project from 2002 to 2008 to address sedimentation and nutrient loading in the lake. Stream bank and channel instability are also of concern. The Oklahoma NRCS prioritized conservation in this watershed. The research focuses on environmental impacts of adoption of conservation tillage systems, exclusion of cattle from streams, and establishment of riparian buffers, and addresses impacts of multi-year climate and extreme climate events.	Spatial - characterization layer with appropriate metadata	Watershed boundary	Q
		Digital elevation model	Q
		Stream network	Q
		Soils	Q
		Geology	Q
		Land use 2005	Q
		Impoundments	Q
		Channel stability	Q
		Conservation practices	I
	Spatial - measurement site layer with appropriate metadata	Weather stations	Q
		Stream gauges	Q
		Soil sampling sites	Q
		Monitoring wells	I
		Characterization wells	I
	Observation - time series	Weather	Q
		Stream flow	Q
		Water quality - biweekly measure of nitrogen, phosphorus, sediment, and geophysical parameters	I
	Observation - ad hoc	Soil quality on farmer fields	I
		Soil physical properties at weather station sites	Q
		Sediment source analysis	I
	Observation - event based	Water quality at gauge sites - six high and six low flow events per year	I
	Survey	Agricultural management	I
		Economic	I
	Model validation file cluster	Hydrologic validation (Starks and Moriasi, 2009)	I

[a] S = on STEWARDS website, Q = in QA/QC by data provider or STEWARDS operations team, I = investigator is collecting or analyzing data prior to publication in STEWARDS, and L = legacy data not yet prepared for STEWARDS publication.

[b] HUC = Hydrologic Unit Code as defined by the U.S. Geologic Survey.

ports (Business Objects SA, San Jose, Cal., and Paris, France), which extracts the necessary information from the database. The use of Crystal Reports for the site descriptions complements reports generated for the methods catalog (described below) and allows the programming for report generation to be within a single code base.

ARS Methods Catalog

In a scientific database, information about the methodology is extremely important. The intent of managing methods information or metadata is to provide adequate detail to allow researchers who were not involved in collecting the data to determine if the data are suited to their intended analyses. Additionally, detailed description of methods provides an important opportunity for researchers to learn from other research groups.

The ARS Methods Catalog is a database to facilitate consistent descriptors of ARS watershed research methods and document field and laboratory methods for measurement data. The ARS Methods Catalog structure is based on the National Environmental Methods Index (NEMI; Keith et al., 2005; NWQMC, 2008) with several modifications for ARS use. The data contained in the catalog allow for a complete description of the constituent measured, the field methods used to collect the samples, and the laboratory methods used to process the samples and derive the measurement value (Sadler et al., 2008). The information in the ARS Methods Catalog is a critical element in the documentation, or metadata, for the measurement data in the STEWARDS database. An example of the types of data provided for field and laboratory methods (table 2) illustrates that many of the fields are not populated for particular methods. In this example, the data were collected from 1979 to 1983 by persons no longer working in the project, so some details of methodology are not available, which is typical of much watershed legacy data. Because the NEMI structure was initially developed for laboratory chemical methodology, it is typical that physical or field measurements may not have applicable information for many of the catalog descriptive fields.

The ARS Methods Catalog is in a proof-of-concept stage of its development and as such is subject to modification as watersheds provide types of data not yet submitted, such as economics or biologic data. The catalog must remain dynamic, as watersheds continue to adopt new measurements and methodologies for evolving research objectives. The relevant methods catalog entries accompany each data set downloaded from STEWARDS.

Managing Metadata

A major goal of STEWARDS is to ensure that critical information regarding data holdings are developed, maintained, and made available to users. Basic flows of metadata are from individual CEAP watershed sites to the STEWARDS metadata database, with an automated harvesting process established allowing STEWARDS metadata elements to move to the Geospatial One-Stop (GOS) metadata warehouse (USGS, 2008) to make the information about the data sets available to a wider potential audience (fig. 2). The files pro–

Table 2. Information about field and laboratory methods provided for a particular parameter selected by user, in this case, total nitrogen in water, collected in the Little Washita River watershed of Oklahoma from 1979-1983.

Descriptive Field	Field Method	Laboratory Method
Method ID	OKGRL_FM07	OKGRL_LM351.2
Official name	Automated water sample collected by Chickasha Sampler	TKN by semi-automated block digestion and colorimetry
Media	Water	Water
Method type	Field	Laboratory
Method subcategory	Physical	Inorganic
Method source	USDA	U.S. EPA National Exposure Research Laboratory
Source citation	Field Manual for Research in Agricultural Hydrology (Brakensiek et al., 1979)	U.S. EPA National Exposure Research Laboratory. August 1993... (EPA, 1983)
Method summary	Automated water sample collection by Chickasha Sampler	This is a semi-automated method. A sample is heated in the presence of sulfuric acid...
Instrument	Chickasha Sampler	Colorimeter
Detection limit type	NA	NA
Detection limit note	NA	NA
Scope of application	NA	This method determines total Kjeldahl nitrogen in drinking water, surface waters, and domestic and industrial wastes.
Concentration range	NA	0.1 to 20
Concentration units	NA	mg/L
Interferences	NA	None given
Precision descriptor	NA	Precision and accuracy values were calculated using...
Quality assurance requirements	NA	NA
Sample handling	Samples should be collected in thoroughly clean plastic or glass bottles and should be of sufficient size to ensure a representative sample. Cool to 4°C.	Samples may be preserved by adding 2 mL of concentrated sulfuric acid per liter of sample and storing at 4°C. However, even with preservation...
Maximum holding time	NA	28 days
Link to full method	NA	http://infotrek.er.usgs.gov/pls/apex/f?p=237:38:514333256505826::::P38_METHOD_ID:9626

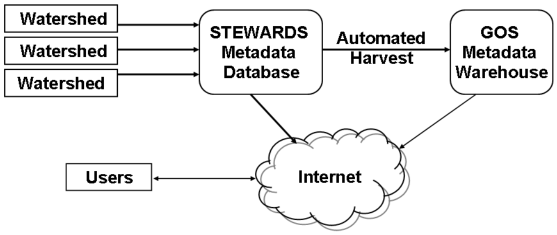


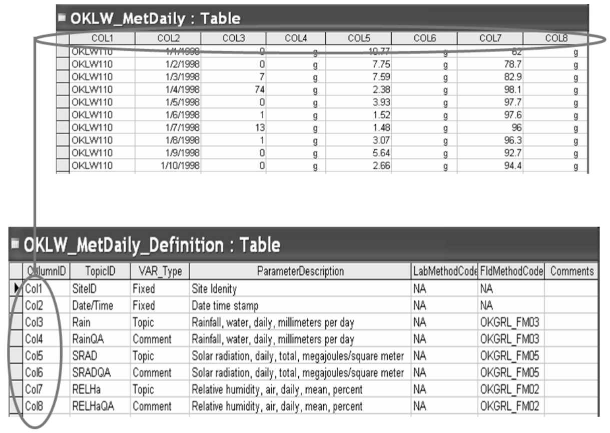
Figure 2. Metadata are provided by watersheds to the STEWARDS metadata database, where they can be automatically harvested for addition to the Geospatial One-Stop metadata warehouse.

vided by watersheds must conform to the Federal Geographic Data Committee's (FGDC) content standard for digital geospatial metadata. STEWARDS users will be able to search and view metadata holdings through either the STEWARDS or GOS website.

Design and Programming Approaches

Site, Spatial, and Temporal Keys

The STEWARDS database is structured to contain data tables and a group of system supportive tables. Each data table contains measurement data for one watershed related to a particular theme (e.g., discharge, weather, water quality). A data_definition table is required to define the contents of each data table. The data_definition table allows STEWARDS to accommodate similar data that are configured into themes differently from different watersheds. For example, one watershed might include sediment load/concentration data in a hydrology table while another watershed might include it in a water quality table. Additionally, one data-data_definition table pair might describe and contain time series data of biophysical measurements, while another might contain economic data, and another might contain survey information. These types of data, and others, can be accommodated within the database design.



OKLW_MetDaily : Table

COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8
OKLW110	1/1/1998	0	g	10.77	g	62	g
OKLW110	1/2/1998	0	g	7.75	g	78.7	g
OKLW110	1/3/1998	7	g	7.59	g	82.9	g
OKLW110	1/4/1998	74	g	2.38	g	98.1	g
OKLW110	1/5/1998	0	g	3.93	g	97.7	g
OKLW110	1/6/1998	1	g	1.52	g	97.6	g
OKLW110	1/7/1998	13	g	1.48	g	96	g
OKLW110	1/8/1998	1	g	3.07	g	96.3	g
OKLW110	1/9/1998	0	g	5.64	g	92.7	g
OKLW110	1/10/1998	0	g	2.66	g	94.4	g

OKLW_MetDaily_Definition : Table

ColumnID	TopicID	VAR_Type	ParameterDescription	LabMethodCode	FldMethodCode	Comments
Col1	SiteID	Fixed	Site Idenity	NA	NA	
Col2	Date/Time	Fixed	Date time stamp	NA	NA	
Col3	Rain	Topic	Rainfall, water, daily, millimeters per day	NA	OKGRL_FM03	
Col4	RainQA	Comment	Rainfall, water, daily, millimeters per day	NA	OKGRL_FM03	
Col5	SRAD	Topic	Solar radiation, daily, total, megajoules/square meter	NA	OKGRL_FM05	
Col6	SRADQA	Comment	Solar radiation, daily, total, megajoules/square meter	NA	OKGRL_FM05	
Col7	RELHa	Topic	Relative humidity, air, daily, mean, percent	NA	OKGRL_FM02	
Col8	RELHaQA	Comment	Relative humidity, air, daily, mean, percent	NA	OKGRL_FM02	

Figure 3. Data table and paired data_definition table prepared by one watershed for upload to STEWARDS in Microsoft Access format.

The use of a generic column ID allows the database to maintain source data from heterogeneous tables with the same base data structure (fig. 3). Each data table starts with COL1 as SiteID and COL2 as the date/time stamp. The rest (COL3 to COL*n*) contain watershed-defined fields for measurement or descriptive data. Each watershed team works from its original data structure and prepares the files for upload to STEWARDS in the required formats (more detail below).

Each watershed team provides GIS layers and associated metadata files to describe the watershed boundary, digital elevation models, and stream networks within the watershed boundary, and other spatial data that vary by watershed. Additionally, each data table that contains data measured at particular sites within the watershed requires a GIS file that includes the coordinates for each measurement site, providing a mandatory link between spatial and temporal data in STEWARDS. The relationships between spatial and temporal data are illustrated in figure 4, where a point on a map must

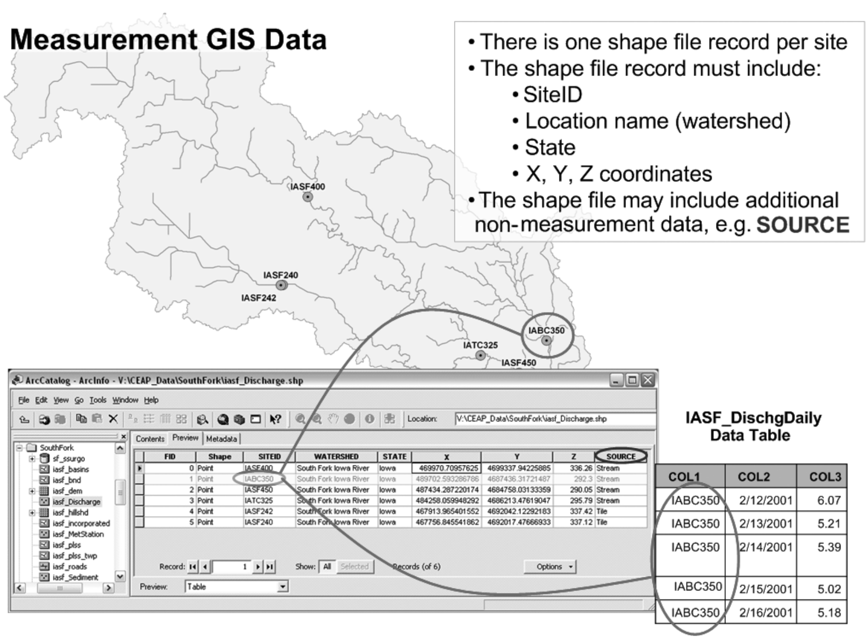


Figure 4. Linkage of spatial and measurement data using the SiteID primary key.

be described in a GIS shape file that provides the SiteID, watershed name, and coordinates. The SiteID in the spatial file is a primary key that links to temporal data (e.g., daily discharge) at a particular site.

System Tables

System tables support the system interface and operations. The system supportive tables include ARS_Locations, ARS_Sites, ARS_SourceTable, Site_Summary, and the ARS Methods Catalog (ARS_ParameterTable, ARS_Methods, and ARS_Analytes tables). The MethodID is used as a primary key, linking between the Data_Table_Definition file and the methods database. The ARS_Analytes table captures information describing the constituents whose value is provided using a method found in the ARS_Methods table. There must be at least one analyte record associated with each method; the method must report the results from a measurement. The one-to-many cardinality between the ARS_Methods table and the ARS_Analytes table is based on the MethodID as the primary key. An additional table, the ARS_ParameterTable, maintains a listing of measurement parameters as a collection of parameter_descriptions from all watersheds. The purpose of the parameter_description is to describe the measurement field in a consistent format. Sadler et al. (2008) provides examples of various types of parameter_descriptions with and without the optional elements. At its minimum, the parameter_description contains the name of the constituent measured, the media, and the reporting units for the sample value. The parameter_description is the primary search field for the STEWARDS interface.

Relationship of Tables

Relationships across data and system tables are core to the STEWARDS software design. As explained above, in the data_definition table, a generic column ID (COL1, COL2, ..., COL*n*) is matched to the appropriate parameter_description. The data_definition table also links the LabMethodCode and FldMethodCode for each parameter to the MethodID in ARS_Methods. Another table that supports queries is the Site_Summary table, which contains the SiteID, TableName, parameter_description, and other information (fig. 5). The Site_Summary table and ARS_ParameterTables are automatically created within STEWARDS as data are uploaded by retrieving the TopicID and associated data table name from the data_definition table. Within the STEWARDS application, the sequence is:

1. Obtain the definition table name by finding all user table names in the system objects of the database.
2. Obtain the corresponding data table name based on the table naming convention.
3. Loop on each data_definition table and perform the next two steps.
4. Obtain the distinct SiteID, BeginDate, and EndDate.
5. Save the SiteID, TableName, parameter_description, BeginDate, EndDate, etc., into the Site_Summary table.

User Interface and User Class

There are three major interface components in STEWARDS: the system entry menu, site-specific and parameter-specific searches, and data access. The user interface was constructed using Microsoft ASP.NET AJAX (Asynchronous JavaScript and XML) to connect the client side (web pages) to the server controls. The server controls includes ArcGIS Server, ASP .NET, Gigasoft ProEssentials (Gigasoft, Inc., Keller, Tex.), and Crystal Decision. Floating Panel web controls were used to create panels on the entry page. Under each panel, menus are created sequentially. For example, under parameter-specific search, a parameter dropdown list, source table list, site menu list box, begin/end date text box, and calendar control are sequentially created, and the parameter selection options are selected in the dropdown list/text box.

The system control interface (entry/menu page) is the gateway for users to access STEWARDS. It consists of the system logo headers at the top, menu buttons, access to maps for watershed/layers selection with zoom and pan functions, and access to system searches. The entry menu also provides user registration and authentication (login optional for viewing). Within the STEWARDS application, publishing a map on the web requires: (1) creating a map using ArcMap, (2) ex-

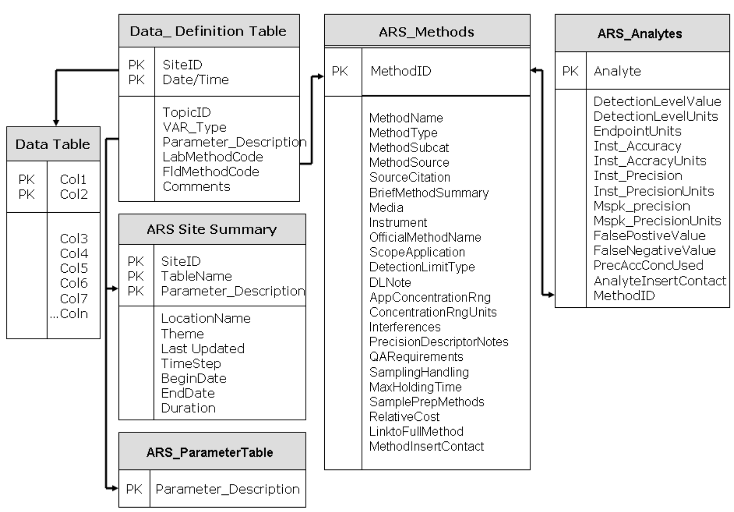


Figure 5. Data structure of STEWARDS showing the relationship across tables. PK represents a primary key that is used to relate one table to another. Each watershed has multiple data and data_definition tables, as needed, to represent the data themes for that watershed. The other tables are system tables, common across all watersheds.

porting spatial data to ArcSDE, and (3) publishing the map document using ArcCatalog.

A user can search using keywords to identify which watersheds and sites provide a particular type of measurement, such as nitrate. After selection of a particular watershed, the user can view the watershed summary or select a site within the watershed either from the GIS map or from the "site-specific search" submenu. For a site-specific search, the user is allowed to access data within a single watershed. After the user selects a site, the Site_Summary table provides the source tables (e.g., Discharge, Weather, Water Quality) available for that site. After selection of one of the source tables, the data_definition table provides the parameter_descriptions available in that table. Once a parameter of interest is selected, the user defines the beginning and ending date, and the data are retrieved based on the selection criteria. Alternatively, for a given watershed, a user can conduct a parameter-specific search. Once the parameter is selected from the menu, a drop-down list, showing the sites where that parameter is measured, is presented for selection. Sadler et al. (2008) described in detail the menu options and steps for conducting searches, visualizations, and downloads.

Users accessing STEWARDS are categorized into three levels with different levels of data access and security requirements. Level 1 users (public access) can search, access, and view data using a web browser and are not required to register. Level 2 users (download access) can perform the level 1 functions plus data download. Level 2 users are required to register so that the data providers can be notified of who is downloading the data that they provided, and users may be notified of updates to the data sets. Level 3 users (password-protected access) may access restricted data. Level 3 users will be USDA employees who have undergone training on allowable use of proprietary or confidential data (level 3 is to be implemented in a future version; to date, no proprietary data have been uploaded).

Implementation of the Operational System

The STEWARDS system is housed at the USDA-ARS National Soil Tilth Laboratory in Ames, Iowa. This laboratory provides operation and maintenance, system administration, user services, data management, and other services. The operational team consists of persons closely linked with the three development components who work directly with the watershed sites to provide training and data screening, conduct QA procedures, and manage the overall progress of the system.

STEWARDS is a customized ArcGIS Server version 9.2 (ESRI Corp., Redlands, Cal.) application running on a Microsoft Windows Server 2003 platform (Microsoft Corp., Redmond, Wash.). The system utilizes Microsoft Internet Information Services (IIS) as a web server as well as the application server. ArcGIS Server 9.2 and Microsoft Visual Studio 2005 were used to create the map-based user interface that is the primary access point to the STEWARDS database. Data visualization displays were developed using Gigasoft ProEssentials (Gigasoft, Inc., Keller, Tex.).

The system configuration (fig. 6) includes an FTP server, a server containing the application and data, map services, and a metadata server (application environment). User requests are received via the web server (internet environment). Requests for standard map operations are forwarded to the map server, which is connected to the database server through a data server. Requests for data are forwarded to the SQL (sequential query language) server for data access. Application programming tools including Microsoft Visual Studio .NET graphics package and SQL were used to develop the user interfaces and other tools. In addition, there are support servers for development and data preparation for upload, and a fail-over environment that replicates the system in the application environment. The fail-over system provides a backup of the system and provides user access in case of loss of service (network, power, security breach) to the primary application environment.

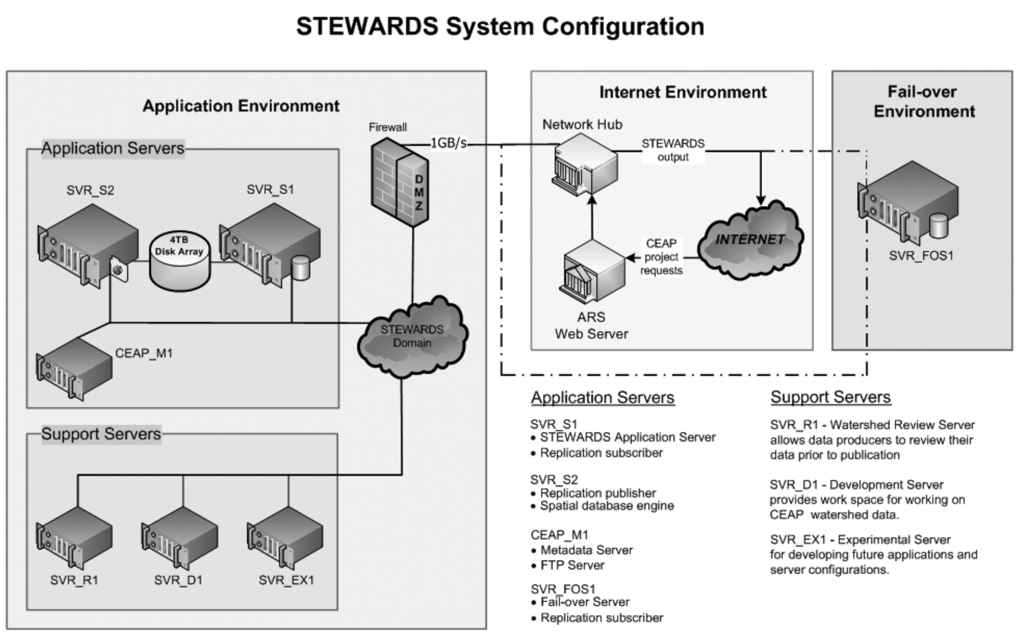


Figure 6. Schematic of STEWARDS data center system configuration, including STEWARDS application environment at the National Soil Tilth Laboratory in Ames, Iowa.

Security is a critical concern in any open-use web application design project. Isolation from the internal user network was a key concern in securing a system like STEWARDS from intruders. The primary and fail-over application environments achieved isolation using both firewalls and configured switches. Once the systems were isolated from production networks, additional measures were taken to protect the systems serving out STEWARDS. Firewalls isolate different subsystems and limit activity to particular servers within subsystems. Additionally, firewalls limit which servers and which ports allow access to outside users. Securing the data was also of concern in the system design. By allowing outside users into the web application server SVR_S1 only, while data are housed on SVR_S2 (no outside direct access), data protection was achieved. Of course, the STEWARDS security design assumes that typical and basic security measures are taken on all subsystems, such as keeping operating systems up to date, updating security, and elimination of open share. If security is breached, the fail-over systems will take over operation until the primary systems are once again secure.

Data Preparation and Upload

The STEWARDS database was constructed by importing Microsoft Access databases from the watersheds (additional detail given below) into the STEWARDS (SQL) database using Data Transformation Services (a data import/export user interface) of SQL 2005 Enterprise Manager. System supportive tables (e.g., Site_Summary table) are created using the Microsoft SQL 2005 Server Tools – Enterprise Manager.

Each watershed team prepares measurement data for themes specific to its watershed, with many themes common across watersheds. The decision was made early in the development process that the watershed teams hold primary responsibility for their data. Given that each watershed team had a functional system for managing local data, there is no requirement that watersheds convert to a standardized primary data management system at the watershed locations. However, in order to upload the data into STEWARDS, considerable effort is required to ensure that data are accurately translated from the local system to the STEWARDS format. The development team works with each site to convert local measurement data from whatever format it is in (e.g., ASCII files, spreadsheets, or databases) to Access databases in the required format.

The data preparation and delivery process is summarized in figure 7. To establish initial data entry for a watershed, the local researchers and data managers worked with the operations team to complete the upload consisting of:

- Descriptive data: watershed and site descriptions, parameter descriptions, and local methods table.
- MS Access data tables, each paired with a data_definition table.
- Spatial data: GIS shape file or raster datasets, each with FGDC-compliant metadata.

As the watershed team prepares data, they work with support as needed from the operational staff for training, templates, tools, and consultation. Sharepoint provides a shared working environment in which local and operational staff can access common files. When the watershed team has completed data preparation, they upload it to an ftp site, where the operations team accesses it and conducts quality assurance procedures on development servers at the operations site.

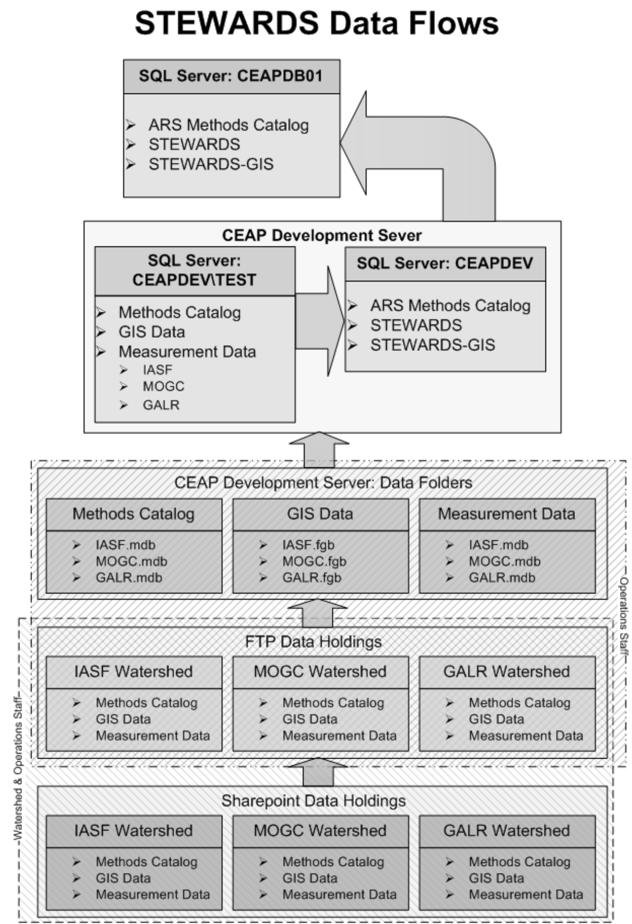


Figure 7. Work flow for data preparation and delivery from the local watershed level to the operational server.

When the data have passed quality tests (often requiring iterative work by the location and operations staff), the data are uploaded into the STEWARDS database.

Discussion

Transition to Operational Phase

The development team achieved the goals of identifying system requirements, developing them into a logical system design, and tailoring the logical data system to a physical data management system that meets the system requirements. The system development and implementation was iterative, whereby an alpha system was tested by the development team in order to improve the system for development of a beta system. The beta system was tested by the development team, as well as other researchers and data managers at the watershed sites, prior to release. The need for iterative database development for complex team research was also emphasized by Scott and Lord (2003) for a multi-site sustainable grazing systems relational database.

Development of STEWARDS involved work at several ARS locations, took five years from the early conceptual design through prototyping to a fully functioning system, and represents one of the primary products of the CEAP watershed assessment studies (Richardson et al., 2008). Project management and resources required during the development process are described by Steiner et al. (2008). Training provided to each watershed team consisted of an overview of STEWARDS, requirements for uploading, and procedures for data preparation. Tools and templates were provided to support preparation of descriptive information, data tables, data_definition tables, parameter names, and entries to the methods catalog. Additional details on data preparation are given by Sadler et al. (2008).

During design and development, the CEAP data management team served as the key decision-making body about priorities for design, content, and structure of STEWARDS. As part of the transition to the operational phase, the developers conducted tests of three critical system development components (website development, software development, and database population) to ensure that the system met the following general criteria:

- Website development: Does the website provide a rich discussion of the project purpose, goals, etc? Are the watersheds and the teams well described? Are there the appropriate links to CEAP NRCS sites, journal articles, and CEAP presentations?
- Software development of the STEWARDS interface: Does the interface meet the functional requirements of the original design process? Does it present the information well and thoroughly, are the output products as intended, and is it easy to use?
- Database population: Are the data sets received from the watersheds complete and in the appropriate format? If not complete, where are the gaps? From a project perspective, is there enough data to publish the database in a meaningful condition?

Access to STEWARDS is provided through the ARS-CEAP Watershed Assessment Studies website (ARS, 2008) to provide the research context and linkage to key articles and related websites. The developers concluded that the database system provides the functions as designed: the ability to navigate across multiple watersheds, data that are spatially and temporally linked on a GIS platform, provision of rich descriptive information about the data, and the ability to search and download data. Because STEWARDS is an ongoing effort, additional features remain to be implemented, such as password protection against access and download of confidential data.

The initial operational STEWARDS had data from seven watersheds relating primarily to weather, hydrology, and water quality. Data providers provided quality-checked data, which is to be updated at annual or other appropriate intervals. There is normally a one or two year time lag between data collection and data publishing in STEWARDS. While the acceptable time lag is not specified, the expectation to publish data in STEWARDS has been included as a deliverable in ARS research plans, and as annual performance criteria for research leaders and scientists at the research watershed locations. The goal is to have five years of data for upload, so the newer CEAP watersheds, started in 2004 or after, are just now approaching the benchmark of five years of data collection, analysis, and publication, prior to posting of the data. One of the significant future challenges of STEWARDS will be the continual preparation and provision of data to the system in order to achieve a comprehensive representation of the ARS watershed research data.

The role of the local watershed teams is critical to understand the local data structures, prepare the data for upload, and minimize the risks of subsequent transcription or data preparation error. Challenges remain in data preparation and upload, both for initial upload from additional watersheds and for periodic updating of data from existing watersheds in STEWARDS. Currently, scientific reward systems offer few incentives to scientists to devote the effort needed to provide data to public databases. Providing such incentives remains a challenge to be addressed by the scientific community in advancing our capacity to conduct complex environmental research that is critical to addressing issues of local, national, and global importance.

The CEAP data management team is evolving into a new management team comprised of STEWARDS developers, the CEAP research community, and CEAP managers who will help evaluate and prioritize future needs, secure internal and partnership resources, and provide oversight of the quality of the service and products provided by STEWARDS.

Unique Features

The STEWARDS developers utilized several unique design elements to meet the system requirements for flexibility, accessibility, and documentation of experimental descriptive information. Because of the inherently spatial nature of watersheds, the decision was made early in the process to implement a GIS-based data system. In addition to typical spatial layers (e.g., watershed boundary, stream network, DEM, soils, land use), there must be a GIS layer for each data table to provide coordinate data for each measurement site. In this manner, the SiteID provides the link between spatial and temporal data, and also provides a link to descriptive information such as methods or site descriptions. The user interface allows the user to move between spatial and temporal views of the data, and to move through links to descriptive information about the watershed, the site, the measurement methodology, and to the measurements themselves.

By designing a generic format whereby each data table is paired with a data_definition table, the developers have provided a data structure that allows one database to accommodate a wide range of data types and formats, and that accommodates changes in the research data within a watershed to meet new research objectives and incorporate evolving research technologies and methodologies. This is an extremely important feature for long-term watershed research, and it allows loosely coupled watersheds, such as in the ARS watershed program, to be presented within the shared framework. As CEAP research efforts expand from cropland watersheds to other types of land use (e.g., rangeland, wetland), STEWARDS offers the capacity to include additional data categories.

STEWARDS provides a wide range of descriptive data, including a watershed description, site descriptions, and methods, in addition to metadata files associated with geospatial information. The descriptive information is provided upon download of data, along with the data policy that specifies how data providers and STEWARDS should be acknowledged in any report or publication using the data. By making the data and information about the data widely available, the data providers also anticipate increased opportunities for collaboration with other watershed research groups.

At the request of the CEAP modeling and risk assessment teams and CEAP National Assessment managers, a feature was added to allow clusters of simulation related files to be archived in STEWARDS. The clusters consist of reports or articles, executable files, input files, and output files associated with key simulation assessments that may provide the basis for future research or that may contribute to policy or programmatic decisions within federal agencies. This provides a record for public transparency, and allows any subsequent analyses to begin with the same information and assumptions that were inherent to the initial work. Such clusters are stored as submitted by the data providers and downloaded in their entirety at the user's request.

Implementation of STEWARDS should provide many benefits by compiling data from multiple research watersheds into a centralized site with well documented methodology and site descriptions. While this article has focused on benefits to the research community, additional benefits to ARS include improved accountability for public funds invested, increased visibility for the national watershed program, increased networking, and capacity building across the research locations. Data access via the internet should be more efficient than when access required contact with one or more individuals at each of the watersheds. In addition, the STEWARDS system provides better long-term security of the data by providing documentation and storage in a central system in addition to the original locations. Additionally, STEWARDS addresses the USDA-ARS's responsibility as a federal agency for transparency and provides the basis for more efficient coordination with other agencies that deal with water (GAO, 2004) and for partnering with watershed research initiatives, such as the Consortium of Universities for the Advancement of Hydrologic Science, Incorporated (CUAHSI), Hydrologic Information System (Maidment et al, 2005).

SUMMARY

The USDA-ARS has developed STEWARDS, a web application used to compile, organize, document, visualize, and deliver time series and geospatial water, soil, management, and socio-economic data for assessment of agricultural conservation practices at watershed scales. The system applies ArcGIS geospatial technologies to provide a flexible approach to visualize and deliver information to the research community. It allows access to the watershed data for internal and external researchers while retaining local control of and responsibility for the data. The system:

- Allows for centralized management of research-quality data from multiple watersheds.
- Interfaces with web-based mapping systems.
- Facilitates tabular data visualization and query functions.
- Enables downloading of research data.
- Compiles, manages, and delivers consistent metadata.
- Supports measurement and quantification of ecosystem effects of conservation practices.
- Facilitates development of policy-planning tools to aid selection and placement of agricultural conservation practices.

Implementation of STEWARDS marks an advance in the research capacity for the ARS research watershed network by providing improved access to well-documented and consistently organized data. Approaches taken during development and implementation of STEWARDS are serving as a model for data management for other ARS research programs and can provide ideas and approaches for use by other watershed research organizations.

This effort represents a move forward for hydrologic and environmental research by providing access to a multitude of data needed to support complex analyses. These data sets represent a significant research watershed network, with many of the watersheds offering the decades of data required to address issues of climate variability and global change (Slaughter, 2000; Slaughter and Richardson, 2000; Harmel et al., 2007). Availability of long-term environmental data greatly enhances our ability to manage water resources under conditions of uncertainty and change (Burt, 2003). Development of STEWARDS required a balance between accountability to a detailed work plan on one hand, and adaptability to meet evolving requirements, resources, and technologies on the other. Major research efforts that extend over many years with numerous organizational players will always require such a balance of accountability and adaptability.

REFERENCES

ARS. 2008. Conservation Effects Assessment Project: Watershed assessment study. Washington, D.C.: USDA-ARS. Available at: http://arsagsoftware.ars.usda.gov/ceap/. Accessed 16 March 2009.

Brakensiek, D. L., H. B. Osborn, and W. J. Rawls, coordinators. 1979. Field manual for research in agricultural hydrology. Agric. Handbook 224. Washington, D.C.: USDA.

Burt, T. P. 2003. Monitoring change in hydrological systems. *Sci. Total Environ.* 310(1-3): 9-16.

Carleton, C. J., R. A. Dahlgren, and K. W. Tate. 2005a. A relational database for the monitoring and analysis of watershed hydrologic functions: I. Database design and pertinent queries. *Computers and Geosci.* 31(4): 393-402.

Carleton, C. J., R. A. Dahlgren, and K. W. Tate. 2005b. A relational database for the monitoring and analysis of watershed hydrologic functions: II. Data manipulation and retrieval programs. *Computers and Geosci.* 31(4): 403-413.

EPA. 1983. Methods for chemical analysis of water and wastes. EPA-600-4-79-020. Washington, D.C.: U.S. Environmental Protection Agency.

GAO. 2004. Watershed management: Better coordination of data collection efforts needed to support key decisions. GAO-04-382 (June 2004). Washington, D.C.: U.S. General Accounting Office.

Harmel, R. D., J. V. Bonta, and C. W. Richardson. 2007. The original USDA-ARS experimental watersheds in Texas and Ohio: Contributions from the past and visions for the future. *Trans. ASABE* 50(5): 1669-1675.

Keith, L. H., H. J. Brass, D. J. Sullivan, J. A. Boiani, and K. T. Alben. 2005. An introduction to the National Environmental Methods Index. *Environ. Sci. and Tech.* 39(8): 173A-176A.

Lane, A. M. J. 1997. The U.K. Environmental Change Network database: An integrated information resource for long-term monitoring and research. *J. Environ. Mgmt.* 51(1): 87-105.

Maidment, D. R., J. Helly, M. Piasecki, P. Kumar, and J. Duncan. 2005. Development of a hydrologic information system for CUAHSI. In *Proc. Watershed Management Conf. on Managing Watersheds for Human and Natural Impacts: Engineering, Ecological, and Economic Challenges*, 607-613. G. E. Moglen, ed. Reston, Va.: ASCE Environmental and Water Resources Institute (EWRI).

NWQMC. 2008. National Environmental Methods Index. Washington, D.C.: National Water Quality Monitoring Council. Available at: www.nemi.gov. Accessed 16 March 2009.

Patino-Gomez, C., D. C. McKinney, and D. R. Maidment. 2007. Sharing water resource data in the binational Rio Grande/Bravo basin. *J. Water Resourc. Planning and Mgmt.* 133(5): 416-426.

Richardson, C. W., D. A. Bucks, and E. J. Sadler. 2008. The conservation effects assessment project benchmark watersheds: Synthesis of preliminary findings. *J. Soil and Water Cons.* 63(6): 590-604.

Sadler, E. J., J. L. Steiner, J.-S. Chen, G. Wilson, J. Ross, T. Oster, D. James, B. Vandenberg K. Cole, and J. Hatfield. 2008. STEWARDS watershed data system: Data development, user interaction, and operations management. *J. Soil and Water Cons.* 63(6): 577-589.

Scott, J. M., and C. J. Lord. 2003. SGS database: Use of relational databases to enhance data management for multi-site experiments. *Australian J. Exp. Agric.* 43(7-8): 729-743.

Slaughter, C. W. 2000. Long-term water data: Wanted? needed? available? *Water Resour. IMPACT* 2(4): 2-5.

Slaughter, C. W., and C. W. Richardson. 2000. Long-term watershed research in USDA Agricultural Research Service. *Water Resourc. IMPACT* 2(4): 28-31.

Simon, A., and L. Klimetz. 2008. Relative magnitudes and sources of sediment in benchmark watersheds of the Conservation Effects Assessment Project. *J. Soil and Water Cons.* 63(6): 504-522.

Starks, P. J., and D. N. Moriasi. 2009. Spatial resolution effect of precipitation data on SWAT calibration and performance: Implications for CEAP. *Trans. ASABE* 52(4): 1171-1180.

Steiner, J. L., E. J. Sadler, J.-S. Chen, G. Wilson, D. James, B. Vandenberg, J. Ross, T. Oster, and K. Cole. 2008. STEWARDS watershed data system: Overview. *J. Soil and Water Cons.* 63(6): 569-576.

USGS. 2008. Geospatial One Stop. U.S. maps and data. Washington, D.C.: U.S. Geological Survey. Available at: http://geodata.gov. Accessed 16 March 2009.

van Evert, F. K., E. J. A. Spaans, S. D. Krieger, J. V. Carlis, and J. M. Baker. 1999a. A database for agroecological research data: I. Data model. *Agron. J.* 91(1): 54-62.

van Evert, F. K., E. J. A. Spaans, S. D. Krieger, J. V. Carlis, and J. M. Baker. 1999b. A database for agroecological research data: II. A relational implementation. *Agron. J.* 91(1): 62-71.