

Sustaining the Earth's Watersheds— Agricultural Research Data System: Overview of development and challenges

J.L. Steiner, E.J. Sadler, J.-S. Chen, G. Wilson, D. James, B. Vandenberg, J. Ross, T. Oster, and K. Cole

Abstract: The USDA Agricultural Research Service has supported watershed research since the 1930s. Data from USDA Agricultural Research Service watersheds have been disseminated independently at each location, hindering multi-site analyses. A virtual team spanning diverse organizational units developed a web-based system, Sustaining the Earth's Watersheds—Agricultural Research Data System (STEWARDS) that allows users to search, visualize, and download soil, water, climate, management, and economic data from Conservation Effects Assessment Project benchmark watersheds. The objective of this paper is to provide an overview of STEWARDS and discuss challenges that were met to deliver STEWARDS on time, according to requirements, and within available resources. The information technology specialists had to understand that vague and changing requirements are reasonable for a system to support loosely coupled research across diverse watersheds. Researchers and data managers had to learn to communicate clearly about their data. Open communication, respect for perspectives and constraints of others, and a shared commitment to the goal provided the basis for trust. Anticipated benefits of STEWARDS include data preservation, increased data use, and facilitation of hydrological research.

Key words: Conservation Effects Assessment Project (CEAP)—hydrologic data—hydroinformatics—informatics project management—Sustaining the Earth's Watersheds—Agricultural Research Data System (STEWARDS)

The need for high-quality, long-term data records of hydrologic systems to address issues such as future water scarcity and potential implications of climate change has been emphasized in major scientific reports (National Research Council 2001; Hornberger et al. 2001). Kinzig et al. (2000) and Kinzig (2001) highlighted the need for increased interdisciplinary research in the area of communicating scientific information, emphasizing potential benefits of information technology (IT) on flows of scientific information to diverse citizen and stakeholder groups. There is a growing international recognition that use of research data is only maximized when data access, management, and preservation are addressed as an inherent part of the research process, and that publicly funded research should be openly available (Arzberger et al. 2004). The National Research Council (2003) report

“Frontiers in Agricultural Research” identified integration of leading-edge science concepts and techniques, of which informatics was an example, as an opportunity for USDA research to better address societal needs. In the face of this growing demand for a new approach to environmental data management, research organizations and teams are grappling with how to organize and support efforts to provide access to complex watershed-scale data.

The Long Term Ecological Research Network (2005) has developed information technologies for synthesis of results across disciplines (Baker et al. 2000), including development of metadata systems (Porter and Brunt 2001) to compile descriptive information to enable researchers who were not involved in collecting or processing the data to understand the details of how the data were collected and processed.

Developing environmental data libraries requires expertise from the data information and computing disciplines as well as expertise in the ecological and natural resources sciences (Baker et al. 2000). The National Science Foundation (2003) has identified the need for a “suite of critical enabling tools for storing, finding, analyzing, and synthesizing a diverse array of data” to enable study of complex systems. More recently, the Consortium of Universities for the Advancement of Hydrologic Science, Incorporated initiated the Hydrologic Information System, “a geographically distributed network of hydrologic data sources and functions that are integrated using web services so that they function as a connected whole.” (Maidment et al. 2005; Consortium of Universities for the Advancement of Hydrologic Science Incorporated 2008)

The USDA Agricultural Research Service (ARS) have conducted watershed research since early in the 20th century with additional sites added periodically to meet evolving research needs. The ARS watershed data have been managed to address location-specific research needs. Data are primarily disseminated independently at each research location, greatly reducing the accessibility of data for multi-site analyses. For a time, rainfall and runoff data were compiled by the USDA ARS Hydrology and Remote Sensing Laboratory (USDA ARS 2003), but after 1990, centralized data compilation and archiving ended. The ARS watershed network as a whole has not implemented many new information technologies, leading to delays and high transaction costs when developing new research thrusts and partnerships such as Conservation Effects Assessment Project (CEAP).

Jean L. Steiner is a supervisory soil scientist, USDA Agricultural Research Service (ARS), El Reno, Oklahoma. **E. John Sadler** is a supervisory soil scientist, USDA ARS, Columbia, Missouri. **Jin-Song Chen** is a former research associate, USDA ARS, El Reno, Oklahoma. **Greg Wilson** is an information technology specialist, USDA ARS, Beltsville, Maryland. **David James** is a watershed specialist for the USDA ARS, Ames, Iowa. **Bruce Vandenberg** is an information technology specialist, USDA ARS, Fort Collins, Colorado. **John Ross** is an information technology specialist, USDA ARS, El Reno, Oklahoma. **Teri Oster** is an information technology specialist, USDA ARS, Columbia, Missouri. **Kevin Cole** is a geographic information specialist, USDA ARS, Ames, Iowa.

In 2002, planning began for building an ARS watershed database to improve efficiency and impact of ARS watershed research investments (Steiner et al. 2003). In 2003, this effort was incorporated into the CEAP, as described by Richardson et al. (2008). The CEAP studies require a variety of data that describe hydrology, soils, climate, topography, management practices, land use, and economics. Such data are also essential for assessments such as the US Environmental Protection Agency's Total Maximum Daily Load program (Turner and Boner 2004), and National Oceanographic and Atmospheric Administration/US Geological Survey (USGS)/USDA's climate and global change program (Cruise et al. 1999).

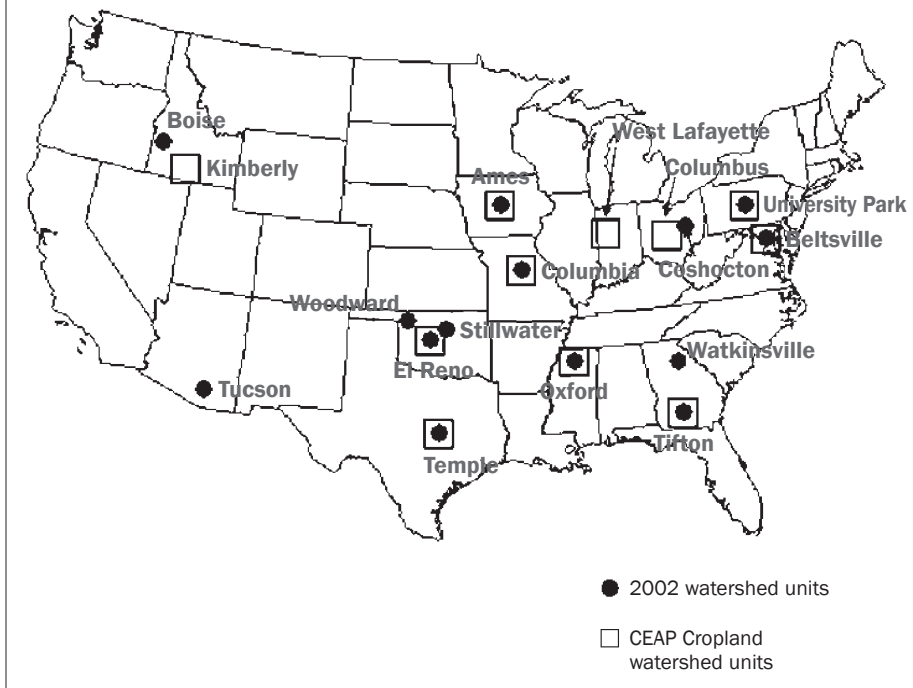
Although a number of research groups are focused on developing cyber-infrastructure for environmental data, the organizational structure and project management are rarely described in the literature. Thus, the first objective of this paper is to provide an overview of Sustaining the Earth's Watersheds—Agricultural Research Data System (STEWARDS), and the second is to describe approaches taken and challenges faced in organizing a virtual team of researchers, data managers, and IT specialists to create and deliver a tool that meets the requirements of the overall CEAP watershed assessment studies as well as other watershed applications of a diverse mix of individual scientists and collaborative research groups. As the STEWARDS database has been considered as a model for other database development efforts, this information may be particularly useful to research data management teams undertaking similar efforts.

Materials and Methods

Watershed Sites. The long-term ARS watershed locations operating in 2002 are shown in figure 1, along with additional ARS watershed research sites that have been developed as part of CEAP. For the pre-2002 sites, the original instrumentation, installation, and data processing procedures for rainfall, discharge and meteorological data were guided by Handbook 224 (Brakensiek et al. 1979—revised from 1962). However, research objectives and data collection evolved differently at each location to address watershed response across varying hydro-climatic regions (e.g., snow, thunderstorm, or groundwater-dominated watersheds). Instruments, parameters observed, and data reduction protocols vary

Figure 1

Locations of long-term USDA Agricultural Research Service (ARS) watershed research units in 2002 and ARS watershed research units conducting Conservation Effects Assessment Project (CEAP) watershed assessment studies for croplands.



among watersheds, although most locations have data about climate, stream discharge, soils, topography, and land management. Data about channel properties and processes are available at some, but not others. Some sites collect groundwater and water quality data, and the parameters monitored vary among sites. Under CEAP Cropland, the focus for the initial release of STEWARDS was on watersheds that have substantial rainfed annual cropping within the watershed, and therefore the initial focus has been primarily on watersheds east of the Rocky Mountains.

For most of the watershed sites, data are not uniformly accessible across the entire period of data collection, nor across different types of data. Climate and hydrologic data are generally most easily available, while land management and vegetative cover records are most difficult to obtain in an easily useable form. Many of the research teams have been addressing these issues but have done so independently of one another (e.g., Hatfield et al. 1999; Slaughter et al. 2002; Sadler et al. 2006; Locke 2004; Bosch et al. 2007; Moran et al. 2008).

Managing a Cross-Cutting, Virtual Team. There is no organizational unit within ARS that has the necessary knowledge, skill set, or mandate to develop a data management sys-

tem to support watershed research. Therefore, we operated as a virtual team, with members from numerous locations within ARS, and with different primary responsibilities in the agency working jointly to address the common goal. From the beginning, the intent was to include researchers and data managers from multiple watershed sites as well as IT professionals. Numerous individuals with expertise in data management in other organizations (e.g., USDA Natural Resources Conservation Service [NRCS], US Geological Survey, NASA, and university faculty) were consulted or participated at various stages of the project.

The team was responsible for developing creative approaches and partnerships to obtain the needed resources. A broad five-year work plan was outlined during a face-to-face meeting of the team co-chairs, a manager in USDA ARS's Office of the Chief Information Officer (OCIO) who later provided staff support to various aspects of the project, a research associate who was fully dedicated to the data system development, and a watershed data manager. We scheduled a face-to-face team meeting early in the project to get acquainted and make initial work assignments. The plan was completed in more detail with feedback from the

entire team and the research watersheds as part of developing the peer-reviewed CEAP watershed assessment study research plan.

Early in the process, we established a Sharepoint website for posting project documents. This allowed team members to have access to current versions of documents prepared and posted by other team members and to check documents out for viewing and editing. In addition, we conducted regularly scheduled web-based team meetings (monthly to weekly depending on the phase of the work) for presentation of new material, exchange of ideas, demonstration of prototypes, and making many key decisions. We held one or more face-to-face team meetings each year.

Our primary annual meeting was in conjunction with the USDA ARS's annual CEAP research meetings, where we could work as a data management team and also schedule time with other teams, such as the methods team, the model validation team, and the model development team. An additional meeting focused on demonstration of the alpha system to watershed researchers for feedback. Another meeting was conducted to review and finalize materials for training at watershed sites on preparation of data for upload and to make assignments for finalizing STEWARDS documentation. In addition, subsets of the team met on an ad hoc basis—e.g., IT, geographical information system, and programming specialists met as needed to work on graphical user interface concepts and functionality, develop graphical user interface layouts and usage storyboards, and develop system architecture diagrams. Another example is ad hoc meetings of the team leaders with national program staff or OCIO managers to discuss resource requirements, plans for the operational platform, etc.

Systems Requirements Development. The system requirements were developed based on review of other data management systems (Steiner et al. 2003), interviews with potential data system users, an inventory of data compiled from ARS watershed research programs, an inventory of data measurements and methodologies used in CEAP research, and input and output requirements for models to be used in CEAP. The system requirements guided the development of the system architecture and features, operating environment, and design and implementation constraints. Key requirements included

Web-based access, spatial platform, documentation of methods, local watersheds retain control of data, not real-time (data delivered in annual updates with a time lag for quality control, analysis and reporting of the data). The system requirements document is available (USDA ARS 2008).

Science Requirements. Because of the complexity of watershed research and the diversity of issues addressed at varied research sites, flexibility to accommodate many types of data was essential. In addition, the ability to capture and document methodology and historical detail was essential. This was the first time for many of the watersheds that the time series and other data were placed into a geospatial environment. Preparation for upload into STEWARDS was also the first time that standard parameter names for common parameters have been adopted across the USDA ARS watersheds and that units have been converted to standard SI protocol.

One of the first requirements adopted was that primary responsibility for data is retained by each watershed research team. The watershed teams are responsible for determining the data to be collected, the most appropriate methods to be used, and quality assurance. Because the data are collected for research, rather than operational purposes, the data will not be posted into the data system in real time but rather added in annual or semi-annual increments. The scientists responsible for the data will have time for quality assurance, analysis, and preparation of scientific reports and manuscripts prior to preparing the data for upload. However, the scientists are expected to upload the data in a reasonable time frame, and requirements for delivery of data to a central database have been included as a research plan deliverable and as part of scientists' and research leaders' performance requirements (table 1).

Agency and Department Requirements. This project must operate within agency and government policies relating to issues such as standardization, quality control, accessibility, and security. The software packages selected were those licensed by the agency and widely available at the watershed research sites. Metadata must conform to the requirements of the Federal Geographic Data Committee (1998). Basic flows of metadata are from individual watershed sites to the STEWARDS metadata database, with an automated harvesting process established allowing STEWARDS metadata elements

to move to the geospatial one-stop metadata warehouse, where they will be widely accessible. Delivery of STEWARDS better positions USDA ARS to support evolving policies toward more open data, such as that defined by the US Global Climate Research Program (1998) and also to address the need for better coordination among federal agencies that collect watershed data (US General Accounting Office 2004).

Results and Discussion

STEWARDS met design requirements developed interactively with the CEAP research community and program managers in early 2006. STEWARDS consists of (1) a centralized site with Web/SQL/ArcGIS servers and application software, including a database management system and a geospatial data access portal; (2) data, including descriptive data, measurement data, and imagery and spatial data; (3) research watershed teams that are data providers; and (4) diverse potential user community (figure 2). The centralized function is supported by the development team (authors of this paper), by operational staff for the data system at the USDA ARS National Soil Tilth Laboratory in Ames, Iowa, and by the staff in the OCIO, Beltsville, Maryland. Data preparation is managed by staff at each watershed location, working collaboratively with the STEWARDS operations staff during data upload to ensure that the data as presented in STEWARDS conforms to requirements and meets expectations (Sadler et al. 2008). A detailed technical description and a link to STEWARDS are provided USDA ARS. 2008. The data are stored as ESRI-ArcGIS, Microsoft Access, or HTML files and are structured and documented in such a way that would facilitate potential future migration to new software systems.

Table 1 shows the project timeline and stages at which key milestones were reached or key issues decided. Although presented as a linear process, the development of STEWARDS was highly iterative. Prototyping of data systems was conducted to explore options, test ideas, answer key technical issues concerning system architecture, demonstrate proposed functionality, and develop needed skills that were later applied to the STEWARDS system, as requirements became firmer or more detailed. In developing STEWARDS, researchers developed new partnerships within the agency to tap into

Table 1
Timeline of STEWARDS development.

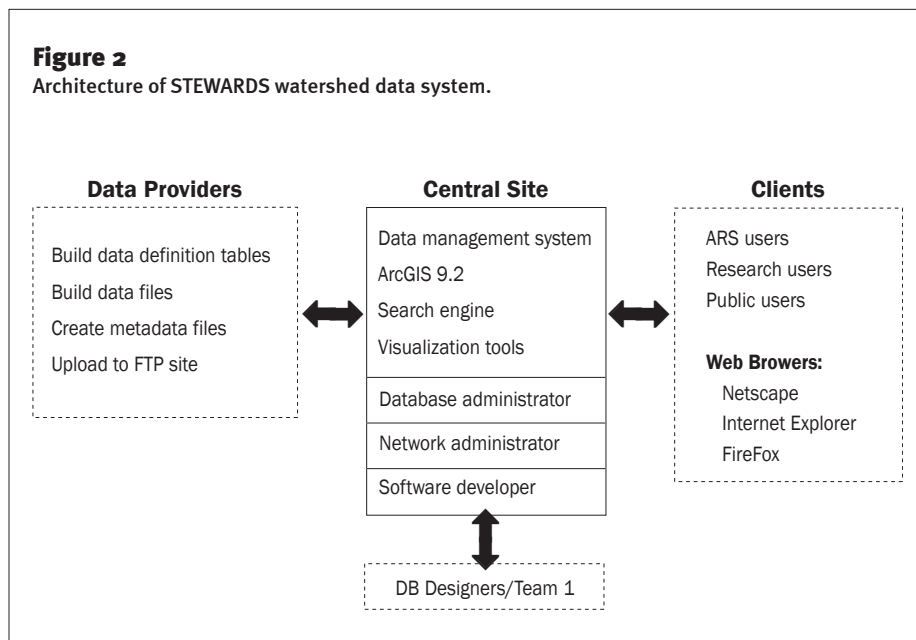
| Phase/task | Timeline | Key players | Key milestones | Key issues resolved |
|--|-------------------|--|---|---|
| Conception and commitment | 6/2002 to 6/2003 | Research leaders National program leaders | Administrator's Research Associate Award. Initiation of CEAP Team 1. | Need for new approach to watershed data management as a national program priority. |
| Develop basic system requirements | 6/2003 to 12/2004 | Research leaders Researchers Data managers GIS specialists IT staff National program leader NRCS IT specialists Watershed teams | Develop plan (as part of CEAP research project). OCIO commitment of staff support. Establish team Sharepoint site and regularly scheduled teleconferences. Inventory of data at watershed sites. | Watersheds retain primary responsibility for data. Must accommodate heterogeneous data. Single point of access with standard formats. Not real time, lagged annual updates. SI units. Web-based. USDA compliant. Browsing does not require password. Download requires password. Data system named STEWARDS. |
| Develop detailed system design requirements | 1/2005 to 12/2005 | Research leaders Researchers Data managers GIS specialists IT staff National program leader NRCS IT specialists Watershed teams | Programmer support from another research project. Data delivery expectations added to research units' charge and to scientist performance plans. | ArcGIS platform. Data table definition for flexibility. Laboratory and field methods. Parameter descriptions. Query by topic. Query within a watershed. Tabular and graphic visualization. Password protected download of proprietary data. |
| Develop system | 1/2006 to 3/2007 | Programmers System operators Data managers GIS specialists IT staff Research leaders National program leader Watershed teams | Prototyping activities for Access databases, spatial display, descriptive files. Commitment for operational support. Develop training material on data preparation. | Laboratory and field methods tables. Parameter naming protocols. Hardware and software configuration. |
| Test and refine system | 7/2006 to 7/2007 | System operators Programmers Researchers Data managers GIS specialists IT staff Watershed teams | Alpha system (ArcGIS 9.1) demonstrated to CEAP researchers for feedback after testing by development team. Beta system (ArcGIS 9.2) demonstrated to development team after test by developers. Training provided at watershed locations on data preparation. Beta system tested by development team and watershed teams. | Schedule for training established and implemented. Schedule for data uploads established. Watersheds with development team members prepared for upload first. Watersheds with longest periods of record trained for next sequence of upload. Newer watersheds to prepare data for upload later. Watershed training and uploads in sequence with most data from earliest trained. |
| Beta release | 7/2007 | System operators Programmers Researchers Data managers GIS specialists IT staff Watershed teams | Beta system installed at operational site. Beta release in conjunction with CEAP Symposium at Soil and Water Conservation Society 2007 Annual Conference. | No proprietary data in version 1.0 that require secure login. |
| Upload additional watersheds and data to STEWARDS | 11/2007 to 9/2008 | System operators Programmers Researchers Data managers GIS specialists IT staff Watershed teams | Complete training and initial data upload at CEAP cropland watersheds. | Implement data delivery schedule. Identify resources to support operational system. |
| Add functions and establish new watershed partnerships | 2008 to 2009 | System operators Programmers Researchers | Establish new system management team. | Evaluate and prioritize needs for new system functionality. Develop and manage new watershed partnerships. Identify resources and team to address new priorities. |

IT expertise. Historically, USDA ARS has maintained a centralized IT staff primarily to support headquarters and administrative functions. Support for IT needs within research programs has been within specific research locations. The development team bridged across organizational segments of USDA ARS to obtain needed skills and expertise. To achieve this, efforts of a “champion” at the national program level and a creative, cohesive team were essential.

Unique Features in STEWARDS. From an IT perspective, it was a challenge to work with a loosely-coupled science group with diverse research objectives and data needs, especially as those objectives and data needs will change over time. To provide flexibility and accommodate diverse types of data, each data table must have a data definition table which fully defines what is located within each column of the data table. The system accommodates various data types including numeric time-series and text. Because watershed data are inherently spatial, data are linked to sites, and because watersheds are inherently variable over time, the time of data collection must be specified.

The SiteID and Datetime fields serve as primary keys for the data system and are the first two data elements provided for each data table. Because data are being compiled from numerous locations, we developed a naming protocol that provides a unique SiteID for each site, consisting of a two-letter state code (IA for Iowa, GA for Georgia, etc.), followed by a two-letter watershed code (SF for South Fork of the Iowa River, WC for Walnut Creek, etc.), followed by a local site ID so that persons familiar with an existing site naming convention in a particular watershed will still be able to understand the SiteID. Each data table requires an associated geographical information system file that links each SiteID to a specific geographical location, providing the key link between the spatial and temporal files.

When the same parameter is measured at different watersheds, the same parameter name is used. Each watershed measuring a given parameter may use different field and laboratory methods, and the methods must be described in the USDA ARS Methods Catalog. The USDA ARS Methods Catalog uses a structure that is compatible with, but separate from, the National Environmental Methods Index (2008). Sadler et al. (2008)



provide additional detail about data file structure and preparation.

At the request of the NRCS CEAP National Assessment Team and the USDA ARS Cropland CEAP modeling and risk assessment team, STEWARDS developed a site to archive and provide download of bundles from key model calibration, or validation runs. The bundles are prepared by the simulation team and should include the executable file of the model version that was used, model calibration and user input files, and model output files. The bundles are compiled as a zip file directory and uploaded into STEWARDS. During download, all files provided by the originator are downloaded for subsequent application.

Training for Data Preparation. Training materials were developed, and training sessions were scheduled at each watershed unit. When possible, one researcher and one person with IT expertise went to each watershed location to communicate the design elements of STEWARDS, benefits to the research community, and procedures to prepare data for upload. The visits provided opportunities for two-way communication, improved the sense of buy-in for the large effort required to prepare data for upload, and increased the development team’s understanding of nuances of the local data management systems that would be involved in the data base population process.

Status. The first version of STEWARDS was completed on time and within resources allocated. The beta version, presented to CEAP researchers in July 2007, provided par-

tial data from three ARS research watersheds that are part the CEAP Watershed Assessment Studies for croplands (table 2). Those selected for the initial upload were sites with members of the CEAP data management team and sites that had long-term data sets at the onset of CEAP. The next version is being developed with the addition of six more research watersheds. Other watershed teams that began research in new watersheds or expanded into larger watersheds at the beginning of CEAP will upload data in subsequent releases. The goal is to have the data from ARS CEAP watersheds entered into the system at the end of the five-year project. Future updates are planned at annual intervals. None of the data uploaded in STEWARDS to date are proprietary in nature. In subsequent uploads, conservation and more detailed management may be uploaded into password protected files that would be only available to USDA employees (or approved collaborators) who were trained in and agreed to meet the terms for use of proprietary information.

The operational team at the National Soil Tilth Laboratory will maintain and operate the STEWARDS data system. The CEAP project homepage website will be maintained through the USDA ARS OCIO. Access to the STEWARDS application will be through a directed link from the CEAP Web site.

Because the development of STEWARDS became integrally linked with CEAP Cropland, several key USDA ARS watersheds are not yet included. Under CEAP Grazing Land, initiated in 2006, several key

Table 2
Schedule for training, data preparation, and delivery from CEAP Cropland watersheds to STEWARDS.

| Research unit location | Watershed | Upload training date | Data in STEWARDS 7/2007 release | Data delivered to STEWARDS 2/2008 | Data in preparation |
|------------------------|------------------------|----------------------|---------------------------------|-----------------------------------|---------------------|
| Ames, IA | Walnut Creek | NA | X | X | X |
| | South Fork Iowa River | | X | X | X |
| Columbia, MO | Goodwater Creek | 3/2007 | X | X | X |
| El Reno, OK | Little Washita River | 2/2007 | | X | X |
| | Fort Cobb Reservoir | | | | X |
| Tifton, GA | Little River | 2/2007 | | X | X |
| Oxford, MS | Goodwin Creek | 3/2007 | | X | X |
| | Beasley Lake | | | | X |
| | Topashaw Creek | | | | X |
| Temple, TX | Leon River | 4/2007 | | X | X |
| Columbus, OH | Upper Big Walnut Creek | 4/2007 | | | X |
| West Lafayette, IN | St. Joseph River | 4/2007 | | | X |
| University Park, PA | Mahatango Creek | 4/2007 | | X | X |
| Kimberly, ID | Upper Snake Rock Creek | 9/2007 | | X | X |

watersheds will be added, including Tucson, Arizona, and Boise, Idaho. In addition, many USDA ARS research projects have legacy data, including a wealth of unit-source watershed data that need to be documented and archived. While some of the legacy data are from sites that still have staff who helped collect the data, as a rule these data may be poorly documented, may not be digital, and key researchers or staff who collected data may no longer be available to assist with understanding the data.

There is a potential to include CEAP USDA NRCS special emphasis watersheds and/or CSREES-funded CEAP watersheds in STEWARDS, but resources would have to be identified to make this practical. As the research at the ARS watersheds continues to evolve, it could be possible to link the data source, STEWARDS, to simulation environments such as the ARS-NRCS-USGS Object Modeling System, in which simulation modules and appropriate data bases could be assembled from a library in order to address specific scientific or natural resources questions. There have been preliminary discussions and may be many future opportunities to link the STEWARDS data system to interactive research environments, such as Consortium of Universities for the Advancement of Hydrologic Science Incorporated Hydrologic Information System or NASA earth sciences research programs.

Challenges. Development of STEWARDS presented many challenges to the researchers, data managers, IT staff, and organizational managers because successful development required working across institutional, disci-

plinary, and technical boundaries and dealing with unfamiliar issues. The scope was potentially so broad, with the original goal “to provide access to data and analytical tools to facilitate multidimensional watershed research and conduct analyses to support individual-, organization-, and policy-level decision making,” that it was essential to define the highest priority and what was feasible. It quickly became apparent that with available resources, the focus would be on data access and that analysis tools would need to reside on the user’s system.

Communication within the team and with other key players was a major challenge. Some of the communication challenges arose from the heterogeneity of the team. Some members approached topics in great detail, others took a broad approach. Most of us focused on issues key to our particular experiences and had more difficulty in focusing on issues outside the scope of our local priorities but that were key to other team members. Terminology was a challenge, with some terms meaning different things to different team members. Sometimes when we thought agreement had been reached, it would later become obvious that additional discussion was required to clarify points of confusion or disagreement. Maintaining an atmosphere of open dialog among all participants and providing a process to resolve issues was essential.

Time constraints were significant. Contrary to recommended best practices for a product development team (Spencer et al. 2006), all members but one had other responsibilities that did not end when they

joined the STEWARDS development team. Our system manager had major responsibilities in agency training and supporting other IT projects. The team leaders both led other research projects, conducted personal research, and accepted details to management positions within the agency during the development process, requiring the other team leader to pick up primary responsibility during those intervals. All of the data managers had significant responsibilities to prepare local data sets and conduct other duties within their units, simultaneously with the team responsibilities. The programmer was responsible for developing a rangelands data system for a different project but worked with this team for the synergistic benefits of gaining geographical information system and other database expertise and research perspectives that influenced the architecture and design of both databases. Other agencies and research groups likely face similar time constraints, but it is highly desirable to have at least some dedicated staff to a major IT development project such as STEWARDS.

Similarly, funding for the project was never secure beyond annual budgets. Within USDA ARS, most resources are committed to salary for permanent employees. Therefore, much of the support was through “in-kind” provision of staff time. Early in the project, the USDA ARS administrator provided partial funding for a research associate, indicating high-level support in the agency for this effort, which was very important in gaining momentum. After establishment of the CEAP project, national program leaders, the CEAP project leader, and research leaders from the CEAP

watersheds provided significant support from CEAP project funds to support the data management team, which allowed the team to develop a more ambitious plan than would have been possible without the CEAP resources and focus.

Because of uncertainty about available resources, trust among the partners was essential. Team managers had to trust team members that they did not support financially nor supervise that they would help define and complete key elements of the project. Those team members had to trust that the project would be seen through to completion and that their roles would be recognized. Research leaders and agency managers had to trust the team leaders that something of value would come back to those who provided staff and financial support to the data management team. Researchers had to trust that STEWARDS would provide benefit to them by improving their efficiency in conducting future research and developing future partnerships and also that they would receive scientific credit for preparing their data for public access in a centralized system. Open communication, respect for the perspectives and constraints of other participants, and a shared commitment to the mission of the team provided the basis for this trust.

There are still many uncertainties about the nature of data needed to support the range of studies that are relevant to CEAP. While the CEAP project plan specifies that economic data are to be included, no watershed to date has prepared economic data for upload. Similarly, biological data, which might include much more text rather than numeric data, have not yet been prepared. However, the system has been designed to accommodate diverse data, and the system developers will work with watershed sites as they prepare some of the additional types of data in future uploads.

There is also ambiguity about STEWARDS as a core research infrastructure that provides documentation, access, and archiving of long-term hydrologic and ecologic data, as an institutional responsibility (important component of the original concept) versus STEWARDS as a system to support USDA ARS's CEAP watershed assessment studies, which focus on particular sites and particular data relevant to CEAP research objectives. It was essential to develop a system that could fill both needs. However, the focus has been to support goals of the CEAP project.

Challenges remain in developing the core research data infrastructure with dedicated long-term resources to support the ARS watershed research program.

Some of the IT team members were unfamiliar with hydrologic and environmental sciences. They were faced with the challenge of designing a system from data sets where the researchers couldn't tell them what might be required for future research needs within existing watersheds nor by future watershed partners. Similarly, the researchers were unfamiliar with the requirements of managing an IT development project, so the IT staff had to find ways to describe what was needed to support the system design and development in a way that was clear to researchers who had to provide much of the information and make many of the requirements decisions. All of the team had trouble determining the types of data needs that might become important when the data were made available publicly, rather than being retained primarily for research applications. The researchers had to grapple with how to adequately describe data for use by those who were not involved in collecting the data.

Summary and Conclusions

Development and release of the STEWARDS data system represents a significant step in developing USDA ARS's organizational approach to managing research information. To contribute efficiently to complex scientific analyses of hydrologic and environmental systems that impact and are impacted by agriculture, it was essential to develop better tools to describe, organize, archive, and gain access to diverse data from long-term watershed research programs.

The development of STEWARDS will enhance ARS's watershed research program through increased communication across watersheds about the nature of data collected, methods of data collection, and future research plans. There is great potential as STEWARDS is fully populated for increased research efficiency of individual scientists and research teams. The STEWARDS development process has enhanced IT capacity and networking opportunities for data managers at the watershed sites. The approach used in developing STEWARDS can provide guidelines, ideas, and tools that may be useful to many other research groups and teams who will face similar challenges.

The team members and agency managers have a greater appreciation for institutional changes that would enhance data management efforts in the future. It became clear that IT expertise is highly variable across watershed teams, and currently there is no structured training and networking system for IT specialists across watersheds. There is a large IT staff with diverse specialized skill-sets who support ARS's administrative and management needs, but there is no formal institutional mechanism to support and facilitate a research information management system that links that IT expertise to national program level applications. These issues are not unique to ARS.

The STEWARDS focus on providing data access is indicative of ARS's commitment—as researchers, staff, research teams, program leaders, line managers, and an agency—to stewardship of data and information from research supported by American taxpayers. Leaders and researchers in many other ARS program areas, such as long-term cropping systems, soil quality, and integrated agricultural systems have indicated an interest using STEWARDS as a model to develop improved data management for key research programs.

Acknowledgements

The authors gratefully acknowledge the contributions of the many researchers and support staff at the watersheds sites who have operated and maintained these research programs through the years. Their faithful and dedicated service provides the basis upon which this natural resource informatics project is built. The development team gratefully acknowledge the support of Mark Weltz and others on the USDA national program staff, Edward B. Knipling for support through the Administrators Research Associate Program, Pete Lombardo and other managers in USDA ARS's Office of the Chief Information Officer, James Ascough for supporting collaboration on programming of the data system, Jerry Hatfield, director of the National Soil Tilth Laboratory where the operational system is supported, and particularly the staff who prepared data for upload to STEWARDS, without whom the development of STEWARDS as we know it would not have been possible.

Disclaimer

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

References

- Arzberger, P., P. Schroeder, A. Beaulieu, G. Bowker, K. Casey, L. Faaksonen, D. Moorman, P. Uhlir, and P. Wouters. 2004. An international framework to promote access to data. *Science* 303(5665):1777-1778.
- Baker, K.S., B.J. Benson, D.L. Henshaw, D. Blodgett, J.H. Porter, and S.G. Stafford. 2000. Evolution of a multisite network information system: The LTER information management paradigm. *Bioscience* 50(11):963-978.
- Brakensiek, D.L., H.B. Osborn, and W.J. Rawls. 1979. Field Manual for Research in Agricultural Hydrology. USDA Agricultural Handbook 224. Washington, DC: USDA.
- Bosch, D.D., J.M. Sheridan, R.R. Lowrance, R.K. Hubbard, T.C. Strickland, G.W. Feyereisen, and D.G. Sullivan. 2007. Little River Experimental Watershed Database. *Water Resources Research*. 43:W09470. doi:10.1029/2006Wr005844.
- Cruise, J.F., A.S. Limaye, and N. Al-Abed. 1999. Assessment of impacts of climate change on water quality in the southeastern United States. *Journal of the American Water Resources Association* 35(6):1539-1550.
- Consortium of Universities for the Advancement of Hydrologic Science Incorporated. 2008. Hydrologic Information System. Consortium of Universities for the Advancement of Hydrologic Science Incorporated. <http://www.cuahsi.org/his.html>.
- Federal Geographic Data Committee. 1998. FGDC-STD-001-1998. Content Standard for Digital Geospatial Metadata. Washington, DC: Federal Geographic Data Committee.
- Hatfield, J.L., D.B. Jaynes, M.R. Burkart, C.A. Cambaradella, T.B. Moorman, J.H. Prueger, and M.A. Smith. 1999. Water quality in Walnut Creek watershed: Setting and farming practices. *Journal of Environmental Quality* 28(1):11-24.
- Hornberger, G.M., J.D. Aber, J. Bahr, R.C. Bales, K. Beven, E. Foufoula-Georgiou, G. Katul, J.L. Kinter III, R.D. Koster, D.P. Lettenmaier, D. McKnight, K. Miller, K. Mitchell, J.L. Roads, B.R. Scanlon, and E. Smith. 2001. A Plan for a New Science Initiative on the Global Water Cycle. Washington, DC: US Global Change Research Program.
- Kinzig, A.P. 2001. Bridging disciplinary divides to address environmental and intellectual challenges. *Ecosystems* 4(8):709-712.
- Kinzig, A.P., S. Carpenter, M. Dove, G. Heal, S.A. Levin, J. Lubchenco, S. Schneider, and D. Starrett. 2000. Nature and society: An imperative for integrated research. Executive Summary. Tempe, AZ: Arizona State University. <http://www.public.asu.edu/~akinzig/nsfes.pdf>.
- Locke, M.A. 2004. Mississippi Delta management systems evaluation area: Overview of water quality issues on a watershed scale. ACS Symposium Series 877:1-15.
- Long Term Ecological Research Network. 2005. LTER Data Catalog. Long Term Ecological Research Network. <http://www.lternet.edu/data>.
- Maidment, D.R., J. Helly, M. Piasecki, P. Kumar, and J. Duncan. 2005. Development of a hydrologic information system for CUAHSI. *In* Proceedings of the Watershed Management Conference—Managing Watersheds for Human and Natural Impacts: Engineering, Ecological, and Economic Challenges, Williamsburg, VA, July 19-22, 2005, ed. Glenn E. Moglen, 607-613. Reston, VA: Environmental and Water Resources Institute, American Society of Civil Engineers.
- Moran, M.S., W.W. Emmerich, D.C. Goodrich, P. Heilman, C. Holifield Collins, T.O. Keefer, M.A. Nearing, M.H. Nichols, K.G. Renard, R.L. Scott, J.R. Smith, J.J. Stone, C.L. Unkrich, and J.K. Wong. 2008. Preface to special section of Fifty years of research and data collection, USDA-ARS Walnut Gulch Experimental Watershed. *Water Resources Research* 44(5). doi:10.1029/2007WR006083.
- National Environmental Methods Index. 2008. National Environmental Methods Index. www.nemi.gov.
- National Research Council. 2001. Envisioning the Agenda for Water Resources Research in the Twenty-First Century. Water Science and Technology Board, Washington, DC: National Research Council.
- National Research Council. 2003. Frontiers in Agricultural Research: Food, Health, Environment, and Communities. Board on Agriculture and Natural Resources, Washington, DC: National Academies Press.
- National Science Foundation. 2003. Environmental CyberInfrastructure: Tools for the Study of Complex Environmental Systems. AC-REE 1. Arlington, VA: Advisory Committee for Environmental Research and Education, National Science Foundation. http://www.nsf.gov/geo/ere/ereweb/ac-ere/ac-ere1_ECI.pdf.
- Porter, J., and J.W. Brunt. 2001. Making data useful: The history and future of LTER metadata. *The Network Newsletter* 14:2:12-13.
- Richardson, C.W., D.A. Bucks, and E.J. Sadler. 2008. The Conservation Effects Assessment Project benchmark watershed—Synthesis of preliminary findings. *Journal of Soil and Water Conservation* 63(6):590-604.
- Sadler, E.J., R.N. Lerch, E.E. Alberts, and T.L. Oster. 2006. Long-term hydrologic database: Goodwater Creek, Missouri. *In* Proceedings of the 2nd Interagency Conference on Research in the Watersheds, May 16-18, 2006, compiled by D.L. Fowler, 161-168. Otto, NC: Coweeta Hydrologic Laboratory.
- 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. Sustaining the Earth's Watersheds—Agricultural Research Data System: User perspective, operation, and application. *Journal of Soil and Water Conservation* 63(6):577-589.
- Slaughter, C.W., D. Marks, G.N. Flerchinger, S.S. Van Vactor, and M. Burgess. 2001. Thirty-five years of research data collection at the Reynolds Creek Experimental Watershed, Idaho, United States. *Water Resources Research* 37(11):2819-2823.
- Spencer, B., R. Butler, K. Ricker, D. Marcusiu, T. Finhold, I. Foster, and C. Kesselman. 2006. Cyberenvironment Project Management: Lessons Learned. Urbana, IL: University of Illinois at Urbana-Champaign. <http://neesgrid.ncsa.uiuc.edu/documents/CPMLL.pdf>.
- Steiner, J.L., D.C. Goodrich, S. Hardegee, M.R. Burkart, T.C. Strickland, and M.A. Weltz. 2003. Information technology applications in the ARS watershed network. *In* First Interagency Conference on Research in the Watersheds, October 27-30, 2003, ed. K.G. Renard, US Department of Agriculture, Agricultural Research Service.
- Turner, B.G., and M.C. Boner. 2004. Watershed monitoring and modeling and USA regulatory compliance. *Water Science and Technology* 50(11):7-12.
- USDA ARS (Agricultural Research Service). 2003. ARS Water Database. Beltsville, MD: USDA Agricultural Research Service. <http://hydrolab.arsusda.gov/wdc/arswater.html>.
- USDA ARS. 2008. Sustaining the Earth's Watersheds—Agricultural Research Data System (STEWARDS). Washington, DC: USDA Agricultural Research Service. <http://arsagsoftware.ars.usda.gov/stewards/>.
- USDA ARS. 2008. System Requirements Specification for STEWARDS Version 1.0 approved. Washington, DC: USDA Agricultural Research Service. <http://arsagsoftware.ars.usda.gov/stewards/documents/STEWARDS%20SysReqSpecSysDesignVer1.pdf>.
- US General Accounting Office. 2004. Watershed Management: Better Coordination of Data Collection Efforts Needed to Support Key Decisions. GAO-04-382. Washington, DC: US General Accounting Office.
- US Global Change Research Program. 1998. US National Assessment of the Potential Consequences of Climate Variability and Change Data Management for Global Change Research: Policy. Washington, DC: US Global Change Research Program. <http://www.usgcrp.gov/usgcrp/nacc/background/meetings/data-policy.html>.