Archiving and Distributing Three Long-Term Interconnected Geospatial Data Sets

Mary Susan Moran, Barbara S. Hutchinson, Stuart E. Marsh, Mitchel P. McClaran, and Aaryn D. Olsson

***Abstract*—Repeat remote sensing field campaigns at experimental sites result in a valuable set of remote sensing data resources, geographic information system data sets, digitized maps, and tabular data that are tied to specific locations. Archiving and distributing these geospatial data generally become the responsibility of local universities and federal research centers with few resources dedicated to the task. The developments in archiving and distributing geospatial data are presented through the description and discussion of three interconnected case studies of data preservation at government and academic units in southeastern Arizona. The main challenges were associated with data archiving, developing online data distribution systems, interconnecting the data distribution systems, and sustaining the archiving and distribution systems. The case studies present multiple approaches to meeting these challenges within the constraints of government and academic units. Results should guide other local efforts to archive and preserve long-term geospatial data and distribute them online.**

***Index Terms*—Archiving, geographic information systems (GISs), geospatial data, Landsat, long term, online distribution, precipitation, runoff, temperature.**

I. Introduction

OVER the past few decades, groups of scientists have come together in multiple field campaigns to study the optical and microwave remote sensing of hydrological and ecological processes. Such efforts are generally focused on local experimental sites that provide a basic foundation for long-term measurements and are enhanced with intensive short-term data acquisition during satellite overpasses. The results are a comprehensive set of remote sensing data resources, geographic information systems (GISs) data sets, digitized maps, and tabular data that are tied to specific locations (termed geospatial data by Morris and Tuttle [43]). These data are particularly valuable because they offer spatial scales ranging from point-based measurements to regional coverage and temporal scales ranging from instantaneous images to short-term image-coincident data to long-term continuous measurements. These place-based geospatial data sets have exceptional research potential for understanding and predicting ecosystem dynamics [40].

There is a movement by government and academic institutions to encourage archiving and distribution of long-term geospatial data[1] for the benefit of both local scientists and the larger international body of students and researchers [3]. This movement is driven by the desire to make data available for research and development and the fear that irreplaceable geoscience data sets may soon be lost. Nonetheless, the archiving and distribution of data from these field campaigns generally fall to local units, such as university departments or federal research centers, with only minimal funding for data archiving and distribution and no dedicated technical support. As a result, these units face a number of common challenges associated with archiving long-term interconnected geospatial data.

The first challenge facing these local units is to archive the data in an electronic digital format that is in a safe location with sufficient standardized metadata to ensure proper application now and in the future. This generally requires a substantial investment of unit resources and is associated with some immediacy to avoid loss of data and institutional memory. Difficult decisions must be made with unit-wide input to select the data and determine the procedures for the archival process. The next challenge is to provide full and open access to these data with an online distribution system. This is particularly difficult because geospatial data sets are commonly archived on different media housed at different sites with a variety of sharing restrictions, protocols, and metadata. Again, the unit must institute a formal planning process to determine which data will be distributed and to design the distribution database and online interface. The third challenge is to integrate the diverse but related data sets by interconnecting the online data distribution systems. This requires a broad community effort crossing federal and university boundaries and leveraging unit resources for a common goal. That goal is to provide the maximum access to all data across units and disciplines to the largest possible group of users, thus facilitating more complex quantitative analyses of ecosystem dynamics. The fourth challenge is to plan for sustainability of the data archiving and distribution systems, taking into account the uncertain futures of technology, data applications, and unit resources. The sustainability of the data archiving and distribution system ensures the continuity of the

Manuscript received January 31, 2008; revised June 11, 2008. First published November 25, 2008; current version published December 17, 2008. This work was supported in part by the UA Institute for the Study of Planet Earth and in part by the NASA Space Grant Program.

M. S. Moran is with the Southwest Watershed Research Center, Agricultural Research Service, U.S. Department of Agriculture, Tucson, AZ 85719 USA (e-mail: susan.moran@ars.usda.gov).

B. S. Hutchinson is with the College of Agriculture and Life Sciences, The University of Arizona, Tucson, AZ 85721-0036 USA (e-mail: barbarah@ag.arizona.edu).

S. E. Marsh and A. D. Olsson are with the Arizona Remote Sensing Center, Office of Arid Lands Studies, The University of Arizona, Tucson, AZ 85719 USA (e-mail: smarsh@ag.arizona.edu; aaryn@email.arizona.edu).

M. P. McClaran is with the School of Natural Resources, The University of Arizona, Tucson, AZ 85721 USA (e-mail: mcclaran@email.arizona.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2008.2002815

[1]In this case, long term refers to "data collected for 20 years or more at a temporal frequency suitable to monitor natural processes" [40].

long-term geospatial data, which, in turn, directly defines the value of the data. This is an important concept that cannot be overstated.

The approaches for addressing these challenges are of interest to government and academic units around the world facing the daunting task of archiving and distributing long-term interconnected geospatial data. This paper presents three case studies describing unit efforts to archive and distribute geospatial data over a multidecadal period for a specific location. The location is the desert grass and shrub biome in southeastern Arizona encompassing the data-rich experimental sites of The University of Arizona (UA) Santa Rita Experimental Range [(SRER); established in 1902] and the U.S. Department of Agriculture (USDA) Walnut Gulch Experimental Watershed [(WGEW); established in 1953]. The archived data include hundreds of images from satellite and high-altitude aircraft sensors acquired over the past 35 years and extensive GIS data sets and digitized maps which are largely available through the Arizona Regional Image Archive [(ARIA); established in 1997]. Coincident with these images and maps are hundreds of measurements of surface reflectance and temperature, soil moisture, vegetation density, cover and leaf area index, and surface roughness, as well as the atmospheric measurements of optical depth and water vapor content. Continuous records of precipitation, runoff, and meteorological conditions over the past 86 years are also available at select sites throughout SRER and WGEW. These case studies are distinguished by the different approaches taken by government and academic units for data archiving and distribution and the cross-unit cooperation required for system interconnection and sustainability.

The overall goal of this paper is therefore to present developments in archiving and distributing long-term geospatial data at local scale. A review of the government and academic responses to the expressed need for data preservation provides a context for the actions taken by the local units. The three case studies offer approaches to the four main challenges to data preservation at the local scale—archiving, distribution, interconnection, and sustainability—that should guide other ongoing efforts to make long-term geospatial data available for future use.

II. Background

> A new digital divide, or rather chasm, is opening up in the scientific enterprise, and something urgently needs to be done to prevent data from being lost into oblivion. ([57]).

A. Government Response

There is growing concern from a number of sectors about the maintenance, long-term preservation, and dissemination of research data [3], [6], [8], [16], [50], [52]. As early as 1982, an international group was formed as the Consultative Committee for Space Data Systems (CCSDSs) to develop standards for long-term digital data handling and storage in support of space research. Working with the International Standards Organization (ISO), the CCSDS released the Open Archival Information System (OAIS) as a draft recommendation in 1999. The OAIS involves identifying digital data, describing them with enough information to adequately preserve them as a unique and authentic entity (metadata[2]), binding that information together into an identifiable package, and facilitating access to the package through search and retrieval tools. It also provides a model for the submission of a package from the information producer to a secure archive and a distribution process from the archive to the information consumer [33]. Subsequently, a 2002 report assessed four preservation metadata schemes and proposed a framework for standardizing the process for a broad range of preservation projects [50]. However, in spite of these efforts, different processes continue to be utilized in various preservation efforts, including such large-scale digitization initiatives as Google Books and Microsoft Live Books [52].

Demands to preserve scientific data in a digital format have occurred coincident with the movement to establishing standards for preservation metadata. For example, a memorandum from the U.S. Geological Survey Water Resources Division (WRD) amended existing policy to require that unaltered field data be preserved in electronic format.

> It is WRD policy that original hydrologic time-series data field-recorded in digital formats and used in automatic digital computation of hydrologic records be preserved indefinitely in electronic digital format. Use of the electronic digital format is required for efficiency in data storage and retrieval and to ensure that the data will be usable for computation if needed in the future. This policy is effective immediately and applies retroactively to all of water year 1999 and prior years to the extent that the necessary electronic digital data files are available for archiving. . .[The purpose is] to ensure no loss of archived data as a result of deterioration of media, obsolescence of hardware or software, or other factors. The procedures should include regular retrieval and checking of data from archive media and copying to fresh media as required. ([55], pp. 1 and 3).

Similarly, the NASA Planetary Data System noted issues raised by the increasing size of data sets, the need for greater storage capacity, and a means to access data across multiple platforms, particularly online [2]. The Geological Society of America, too, has recognized the fundamental importance of having policies in place that support good practice in preserving and making accessible often irreplaceable geoscience samples and data sets. "Many of these collections, including those housed by federal and state agencies, universities, museums, private companies, and individuals are at risk because of severely limited space and funding for proper curation." ([54], p. 1). Furthermore, a Department of Energy 2002 report on long-term spatial data preservation concludes with the statement that the resolution of this issue will require a combination of technological, management, and policy decisions while at the same time recognizing the following.

> The cost and complexities of moving digital information forward into the future raise our greatest fear about

[2]"Data about data" or here used as records providing data about a data object.

the life of information in the digital future: namely, that owners or custodians who can no longer bear the expense and difficulty will deliberately or inadvertently, through a simple failure to act, destroy the objects without regard to future use ([3], p. 12).

Throughout this period, a number of related reports from the National Academies were published, documenting the potential and real loss of intellectual memory through technological obsolescence and inadequate information handling practices. Moreover, addressed were the difficulties that scientists face when addressing these goals without adequate funding or technical support [45], [46]. One response to these difficulties has been the Canadian-directed International research on Permanent Authentic Records in Electronic Systems project involving a multidisciplinary approach for developing the theoretical knowledge and practical approaches necessary for progress in this area [15]. Recently, an international "Alliance for Permanent Access" was initiated at the "Second International Conference on Permanent Access to the Records of Science" for the purpose of creating a European digital information infrastructure that will develop preservation strategies guaranteeing open access, interoperability with other data sets, and archival repositories [57].

Another more recent and agency-specific effort resulted in a statement of principles for the management and preservation of environmental[3] data emanating from the National Oceanic Atmospheric Administration (NOAA). Again, under the auspices of the National Academies, the principles defined include "full and open access to data" for end users, which provides for "discovery, access, and integration"; resources for end-to-end data support; stewardship for metadata generation; and a formal and ongoing planning process that includes broad community input into the selection of data to archive [47].

Finally, the U.S. National Science Foundation (NSF) has launched several programs through its Office of Cyberinfrastructure "to facilitate the development of new applications, to allow applications to interoperate across institutions and disciplines, to ensure that data and software acquired at great expense are preserved and easily available, and to empower enhanced collaboration over distance, time, and disciplines" [16]. For instance, the Sustainable Digital Data Preservation and Access Network Partners program serves to develop new types of organizations that will develop "new methods, management structures and technologies to manage the diversity, size, and complexity of current and future data sets and data streams" [48].

B. Academic Response

Simultaneous with these government-related efforts, there has been a conversation, particularly among academic scientists, librarians, and archivists, on strategies for the preservation and dissemination of digital materials emanating from universities. In 2002, the Council on Library and Information Resources and the Library of Congress issued a report documenting the potential cultural loss if the problem of preserving digital media was not solved. The solution was tied to a combination of technical, economic, and legal changes placed within the context of innovative policy development [6]. At the same time, the Scholarly Publishing and Academic Resources Coalition published a report making the case for academic institutional repositories to expand access to research data to gain greater control over scholarly communication output. Ultimately, as the number of universities involved in such initiatives increased, the result would be revealed in a collection of interoperable repositories containing a vast corpus of self-archived heterogeneous data that would facilitate interdisciplinary research and discovery [7], [34]. In both cases, libraries would provide a custodial role in managing and disseminating the intellectual output of their institutions. To this end, in 2004, the Library of Congress asked university research libraries to develop a National Digital Information Infrastructure and Preservation Program (NDIIPP) [51]. This program has focused attention on the need for preservation standards and policies and has helped develop tools and services for preservation in cooperation with a network of partners.

The importance of preserving and providing access to data sets has become a more visible topic of discussion, not just within the university library community but throughout the academia in the past several years. The need to build short, medium, and long-term availability of research data into project plans has been strongly recommended to ensure the ability to verify research results, provide for longitudinal and interdisciplinary use of data, and to secure against physical decay of data storage, loss of metadata for retrieving information, and the potential inability to run outdated software [9], [16]. Responses to this need have varied, but there have been a number of universities that have implemented data archiving systems. Notably, the University of Michigan provides a "best practices" fact sheet for producing preservable data sets as part of their library-based "Deep Blue" institutional repository [59], and the University of California Irvine's Department of Information and Computer Science offers the "Knowledge Discovery in Databases Archive" developed through an NSF grant as a permanent repository of publicly accessible data sets for research [24]. This past year, Cornell University's Mann Library received a three-year NSF grant to create a set of services and tools to make it easier to share digital scientific data among academic and government repositories [5].

Supported by the Library of Congress, the National Institutes of Aging, and NSF, Harvard University has initiated the "Dataverse Network" as an infrastructure for hosting and preserving data presented in scientific publications. The Network software is freely available and gives authors and data producers a reliable repository for their data along with a rich metadata capability that allows for more complex statistical analyses [32]. Specific to the issue of geospatial data is the North Carolina Geospatial Data Archiving Project, which is a joint project of the North Carolina State University Libraries and the North Carolina Center for Geographic Information and Analysis. Recently funded by the Library of Congress' NDIIPP program, the project focuses on the "collection and preservation of digital geospatial data resources from state and

[3]Environmental data is defined to include all types of Earth System observations, including remotely sensed data.

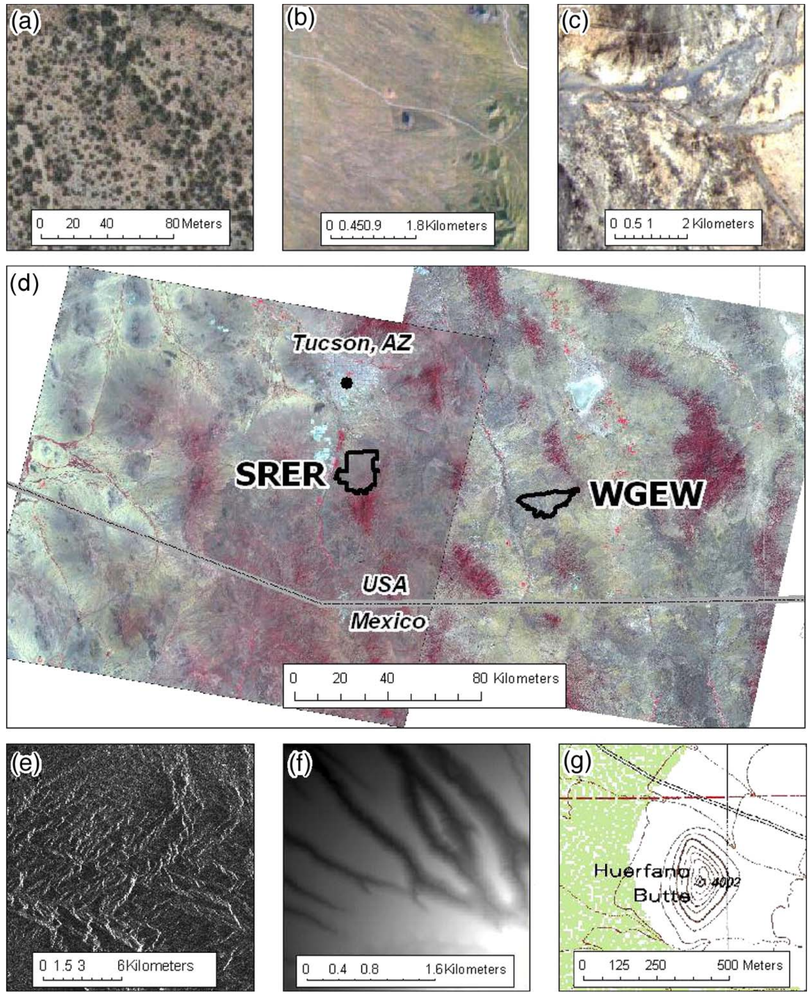

Fig. 1. (d) (Center panel) Location of WGEW and SRER with respect to Tucson, AZ, the U.S.–Mexico border, and the footprints of two Landsat scenes. In addition to the (WRS2 path 36/row 38) 42 Landsat scenes covering SRER and (path 35/row 38) 85 scenes covering WGEW, ARIA contains imagery with high temporal resolution (MODIS and AVHRR; not pictured). (a) High spatial resolution (DOQQ). (b) High spectral resolution (AVIRIS). (c) MTI. (e) ERS-SAR. (f) DEMs. (g) Topographic maps (DRGs).

local government agencies” ([43], p. 1) with an emphasis on workflow and archival processes.

It is worth noting that there has been some effort to combine government and academic responses to the need for data preservation and access. For instance, the newly established EcoTrends project (http://www.ecotrends.info/EcoTrends/index.jsp) involves multiple agencies and institutions to make long-term ecological data available for cross-site data exploration and analysis [17]. An international initiative is underway to improve compatibility, quantify the accuracy of land cover data sets, and develop a common land cover language to interpret map legends [26]. At the same time, a recent announcement from Google disclosed plans for a new initiative to provide a massive storage facility for scientific data sets that will be freely available to scientists and other users alike. Included in this initiative will be Hubble Space Telescope data [35]. However, in spite of the increasing visibility and opportunities for long-term preservation and access to research data, currently, it remains largely up to individual scientists, local agencies, academic units, and libraries/archives to implement the archival and distribution technologies, guidelines, and policies.

III. Location

The case studies are set in southeastern Arizona, encompassing the Chihuahuan and Sonoran Deserts surrounding Tucson, Arizona. Southeast Arizona is characterized by mild to cold winters, rainfall 50%–80% in summer (mid-June to mid-September), and prevailing vegetation of low shrubs and grasslands. The area is covered by a single MODIS scene and by multiple Landsat scenes (Fig. 1). To preserve these remote sensing resources along with a variety of other digital image and map data, the ARIA was created as an interdisciplinary resource. Within this region, there are two important experimental sites—WGEW and SRER—administered by the USDA and UA, respectively.

The 21 500-ha SRER is located 60 km south of Tucson, AZ. SRER was founded in 1902 and administered by the United States Forest Service until 1987 when those responsibilities were assumed by the UA College of Agriculture and Life Sciences (CALS). The elevation ranges from 900 to 1300 m, and the average annual precipitation ranges from 270 to 450 mm along that gradient. The mean annual temperature is approximately 16 °C, with freezing temperatures occurring 20 to 40 nights in the winter, and summer daytime temperatures regularly exceeding 35 °C. It is a semiarid shrub savanna, with stands of grass punctuated by a diversity of shrubs such as mesquite, creosotebush, cacti, and other species [36]. Long-term measurements and repeat photograph collections have been collected in a systematical manner on the SRER since its establishment.

The 15 000-ha WGEW is located at the transition between the Sonoran and Chihuahuan deserts near the town of Tombstone, AZ. WGEW is administered by the USDA Agricultural Research Service (ARS) Southwest Watershed Research Center (SWRC). WGEW elevation ranges from 1220 to 1950 m, with desert shrubs dominating the lower two-thirds of the watershed and desert grasses dominating the upper one-third. The climate at WGEW is classified as semiarid, with mean annual temperature at Tombstone of 17.7 °C and mean annual precipitation of 312 mm. Hydrological data, including rainfall and runoff, have been collected on experimental watersheds operated by the ARS in southern Arizona since the 1950s [53]. Precipitation is measured with a network of 88 weighing-type recording rain gauges arranged in a grid throughout the watershed [23]. Hydrometeorological and soil erosion/sedimentation data are collected from over 125 instrumented installations on WGEW [30], [58]. On a daily basis, all locations are automatically and sequentially queried, and data are transmitted to a dedicated computer at the Tombstone field office. In addition to the long-term high-quality hydrological data collection, hundreds of satellite and aircraft spectral images of WGEW have been acquired [39], and extensive GIS databases are available [25].

IV. Case Studies

Case studies are used here to provide specific examples of how long-term research data sets have been managed in both the government and academic communities. The studies include the following:

1) ARIA maintained by the CALS Arizona Remote Sensing Center (ARSC) and the UA Committee on Remote Sensing and Spatial Analysis (CRSSA);
2) SRER Digital Database (SRER DD) maintained by the UA/CALS;
3) SWRC Data Access Project (DAP) and WGEW Image and Ground Data Archive (WIGDA) maintained by the USDA-ARS SWRC.

A. ARIA

ARIA is an interdisciplinary resource system and online clearinghouse for digital image and map data for the Sonoran Desert region, including the U.S. Southwest and northern Mexico (http://aria.arizona.edu). The focus on regional information provides a direct and tangible link between research and instructional programs in the earth science disciplines, particularly as they relate to climate variability and land cover change. ARIA was created out of a three-year NASA Center of Excellence Grant in 1997. UA Professor Robert Schowengerdt, along with Coprincipal Investigators from the UA CRSSA, developed ARIA with an initial budget of approximately $370 000. Originally designed to facilitate local sharing of data and avoid the cost of making duplicate acquisitions of satellite images, aerial photography, digital elevation models (DEMs), and map data, ARIA has grown to serve over 2.0 TB of data online. Over 14 000 users have downloaded over 500 000 scenes from ARIA since 1998 and currently download approximately 50 scenes per day.

Although over 90% of the downloads are from the users in the U.S., primarily from Arizona and the southwestern states, there are also registered users from 55 countries. These users come from a wide range of backgrounds, including education (33%), federal government (16%), commercial (14%), and state and local governments (13%), as well as nonprofits (1%) and tribal governments (1%). There are over 125 registered educational users. Users from the UA account for 20% of the downloads, Arizona State University 14%, and Northern Arizona University 13%.

ARIA contains over 1400 Landsat multispectral scanner/thematic mapper/enhanced thematic mapper plus (MSS/TM/ETM+) scenes from 152 WRS2 tiles and includes 42 and 85 scenes that cover the SRER and WGEW, respectively. Satellite imagery comprises 33% of the ARIA data holdings. There are also over 6500 digital orthophoto quarter quadrangles (DOQQs) from 1992, 1996, and 2005 and other aerial photography representing 55% of the archive, as well as digital raster graphic (DRG) maps (2%) and DEMs (2%). A variety of other image and map data including NASA airborne visible/infrared imaging spectrometer (AVIRIS), NOAA advanced very high resolution radiometer (AVHRR), NASA-MODIS, NASA-ASTER, DOD multiband thermal imagery (MTI), European Remote Sensing synthetic aperture radar (ERS-SAR), NASA spaceborne imaging radar-C, and SPOT-HRV data comprise the remainder of the archive (see examples in Fig. 1). The selection of data for inclusion in the archive was originally based upon what data sets were available among the researchers at the UA and local, state, and federal organizations in the area. After that initial data mining, additional data sets were acquired opportunistically based upon the availability of data and an assessment of need. In particular, newly available high spatial resolution (1-m) aerial photographs (DOQQs) were added to the database because of a high demand.

Most recently, data for the WGEW were updated, and over 100 Landsat MSS, TM, and ETM+ images and dozens of microwave images were added to ARIA. For ARIA, commercial providers SPOT Image, Space Imaging/EOSAT, and Eurimage agreed to waive the restrictions associated with their license agreements to allow the images to be shared under a reasonable set of conditions defined by each company [41].

ARIA has become an attractive resource for acquiring remotely sensed imagery and GIS layers for research (e.g., [44] and [56]). Researchers have utilized ARIA data sources to publish peer-reviewed journal articles based on AVHRR [29], [61], Landsat [1], [10]–[13], [20]–[22], [27]), DOQQs [49], [60], and DEMs ([14]). This list includes articles in which ARIA was acknowledged; the full list is likely much larger. The original ARIA metadata schema was created in 1997 to accommodate application requirements (e.g., searching, previewing images, and download) and user requirements (e.g., processing history,

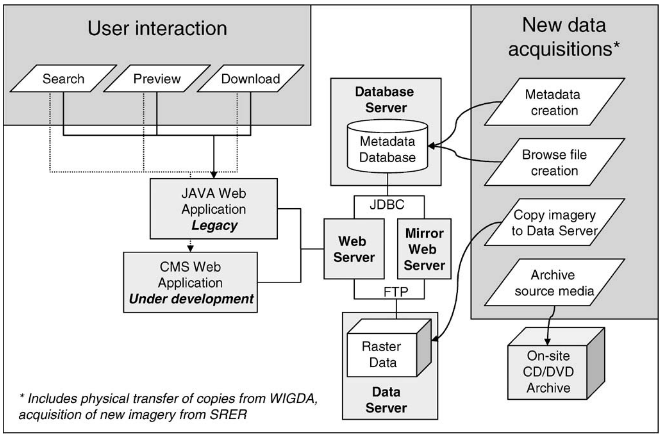

Fig. 2. Architectural diagrams of the ARIA and interconnections with WIGDA and imagery from SRER.

spatial reference, and data organization) based on the Federal Geospatial Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM), which was still being developed at the time of ARIA's conception [18]. The CSDGM did not specify an output format; therefore, the application was developed to output metadata in an HTML table. Since the publication of the CSDGM, the ISO has published an abstract standard, which is the ISO-19115:2003 for describing metadata for geospatial metadata, and an implementation standard, which is the ISO/TS 19139:2007 which describes the formatting of the metadata in XML [19]. This standard will supersede the FGDC's standard, but the FGDC has released a tool for translating the FGDC-STD-001-1998 into ISO 19139, which the American National Standards Institute will adopt as the official metadata standard for geographic data. ARIA developers intend to add functionality to support the ISO 19139 standard by the end of 2008.

At this time, the preservation of the data for long-term access is only provided by copying from the original source onto CD, digital versatile disk (DVD), and ultimately onto redundant array of inexpensive disk (RAID) data storage devices. The metadata for each image and map are created before images are made available online. Metadata includes the ten top level entities, as prescribed by the CSDGM: identification, data quality, spatial data organization, spatial reference, entity and attribute information, distribution information, metadata reference, citation, time period, and contact information, although the format does not pass CSDGM requirements. Browse images at low, medium, and high resolution are also created and are available for viewing prior to downloading.

The ARIA system comprises hardware, e.g., servers and data storage devices, and software, e.g., web application and metadata database (Fig. 2). Since its 1997 inception, every component has been replaced at least once due to additional data requirements, hardware failure, or newly available software, including tools for development. Additional storage was purchased in 2002 and 2007. A new server solution was purchased in 2007 to replace the original server, which is a Sun E-450, which encountered unrecoverable hardware failure in summer of 2007 and was replaced temporarily by an underpowered workstation. The occasional additions to ARIA of capacity or hardware, all of which occurred after the original funding was completed, reflect the goodwill of UA researchers, departments, the Institute for the Study of Planet Earth (ISPE), and the Vice President for Research. In addition to hardware upgrades, the software ARIA runs on has been upgraded numerous times to keep up with currently supported web servers, servlet engines, and relational database management systems (RDBMSs).

The original hardware configuration included two 600-disk CD/DVD jukeboxes, each of which contained at most four readers. Data were stored on CD or DVD and accessed through a web application as near-line storage. When a user requested a specific data granule, the corresponding CD or DVD was mounted, and data were streamed to the user over the web after an 8-s delay to ensure proper mounting of the optical media. In 2002, the two jukeboxes were replaced by a small computer storage interface integrated drive electronics RAID with 3-TB capacity, and an additional 5-TB capacity was added in 2008. The original server (a Sun Enterprise E450) lasted ten years before hardware failure necessitated its replacement. An interim solution had ARIA running on a workstation until new hardware could be purchased and ARIA deployed on it. The new solution includes separate servers for the web application and database, each with a quad-core Xeon processor.

The web application was written as a set of Java servlets (version 1.1) which ran under Jrun 2.2 and Netscape Enterprise Webserver. In 2002, ARIA was migrated to Apache webserver 2.0 and Apache Tomcat 5.0.28 servlet engine running Java 1.4, and the entire application refactored to run as a Struts application. In 2008, ARIA was deployed on a new server running Java 1.6 and Tomcat 6.0.14, which uses servlet version 1.5. The web application uses a metadata database to allow users to search the archive and download data. Specifically, it links the physical location of the data to information about the scene itself. The original metadata database ran on Interbase 1.5 RDBMS. In 2002, the database was migrated to Oracle 9.1 and, in 2004, to MySQL 5. These migrations reflect gradual increasing maintenance costs of legacy systems. Updating ARIA software and refactoring the code have raised the baseline for legacy maintenance and will benefit current and future system managers as they deal with an online resource with more than ten years of history.

The process of adding new data to ARIA has not drastically changed over time. New scenes must be accompanied by a CSDGM-compliant metadata record that fully describes the granule and a set of browse images of small, medium, and large sizes. The metadata include information about their spatial extent and resolution, the nature of the data (e.g., satellite imagery, maps, DEMs, and aerial photos have quite different metadata requirements), the source of the data, current owner, usage constraints, and metadata author. While some data were acquired with a full and detailed metadata record, most did not. The variety of data formats, which are metadata formats, has required a diverse set of tools to be used for deriving metadata for what is an involved process of metadata authorship. In many cases, metadata must be acquired externally, for example, to determine the acquisition date of DOQQs, the sun angle of a satellite image, or to identify a projection or datum given a numeric code. In other cases, metadata must be extracted through a variety of means from proprietary software (e.g., identifying the spatial reference information including projection datum and minimum bounding rectangle). The specific software used varied, depending on the data format and experience of the authors. Over the years, 15 undergraduate and graduate students, as well as ARSC staff, have entered metadata records and created browse images for ARIA. Image processing software used to derive spatial reference and data organization information includes Leica Geosystems Erdas Imagine, Environmental Systems Research Institute (ESRI) ArcGIS, Environment for Visualizing Images, and Photoshop. Perl, Java, and AML scripts have been developed to facilitate entry for common data and to ingest metadata of different formats.

After the initial development of ARIA, and when both NASA and UA matching funding came to an end, the oversight of ARIA shifted to the UA CRSSA. Direct maintenance became the responsibility of ARSC, and ARIA has remained in existence as a result of both financial contributions from UA remote sensing groups (~$75 000) and the contribution of personnel time by ARSC (~0.25 Full-Time Equivalent Employee System Management). This operating status has allowed ongoing access to ARIA but has been an impediment to real growth because of the lack of dedicated personnel and funds to plan for greater sustainability and growth in the data archive and enhancement of the user interface. A cost-benefit analysis conducted in 2005 indicated that ARIA could begin charging nominal amounts for some of the data sets, and it becomes "profitable" after approximately two years. However, the required initial investment of approximately $50 000 to implement these capabilities has not been found. Nevertheless, internal commitments to ARIA remain strong, and it will continue to be an important resource for the foreseeable future.

B. SRER DD

The SRER DD is an online accessible collection of geographically explicit long-term vegetation and precipitation measurements, and repeat photography images spanning over 100 years (http://cals.arizona.edu/srer). In addition, spatially explicit ancillary data for soils, elevation, and infrastructure are made available to assist in the analysis of the primary data. Regular updates are made to the primary and ancillary data sets.

The digital archive of the primary vegetation, precipitation, and repeat photography data was created by manual entry from paper data sheets and scanning original photographic prints. Prior to digitization, the paper archive was stored in cardboard boxes in the wooden office building at the SRER headquarters. Initially, the UA Library Special Collections was approached to archive this material, but that option was rejected because access to potential users would have been severely limited and cataloging would have taken many years. The final decision to create a digital archive was inspired and hastened by a wildfire that threatened the collection in 1994. At that point, the collection was hurriedly rescued from its remote and vulnerable location and brought to the UA campus, and plans began developing to digitize the collection and make it available on the Internet so that broader access would fulfill the promise and potential of the publicly funded research.

Fig. 3. Images from location 42.03 in the repeat photography collection SRER DD archive. The scenes show the history of a 5-ha experimental thinning and removal of mesquite (*Prosopis velutina*) performed in 1945. (Top left) In June 1936, mesquite cover was continuous at the base of the distant Santa Rita Mountains; (top right) the sharp corners of the mesquite removal are obvious in the midground in June 1956. (Bottom left) By January 1986, mesquite abundance had increased in the 5-ha experimental area, and ocotillo (*Fouqueria splendens*) became established in the foreground; (bottom right) by April 2007, mesquite cover had fully recovered from the removal experiment performed 60 years earlier.

The continuity and length of record were the primary criteria used to prioritize the archiving of greater than 100 different data sets in the entire paper archive. Those data sets with recent measurements and a long history of repeated measurements were the highest priority for creating the digital archive [36]. These data sets include four long-term (as early as 1956) studies recording vegetation measurements in response to livestock and mesquite removal practices. Every three years, repeated measurements are made at these 130 locations and are updated to the Web site. Data from the monthly precipitation measurement at 24 rain gauges (13 of which began in 1922) were archived, and new measurements are regularly updated. Images and associated text from 110 repeat photography locations (many started in 1902) were also included, and they are updated on a regular basis (Fig. 3). Finally, an annotated bibliography [37] of over 450 references by more than 200 authors was archived as hyperlinked text, and updates are maintained for currency.

In the near future, additional historical information including study plans, data, maps (.pdf format), and photographs (.tiff) will be digitally archived for 80 research locations that were active from 1903 to 1940. They will be searchable by keyword, and organized and presented in linkable tables and maps.

The supporting spatial database information, such as digital elevation, topography, soils, roads, fences, and water developments, was archived to assist with the analyses of the vegetation, precipitation, and repeat photography data sets. In addition, the Universal Transverse Mercator (UTM) coordinates of all vegetation measurement locations, rain gauges, and repeat photography locations are included in the archive.

Vegetation abundance measurements include plant density, cover, and biomass measured at permanent transect locations. These data are provided in spreadsheet format (.xls), and associated study plan descriptions and metadata are available as text format (.txt). Monthly precipitation measurements are also provided in spreadsheet format (.xls), and associated location histories and metadata are given as text format (.txt). Repeat photography images (.tiff) and associated text (.txt) are searchable by keyword, can be accessed from a table, or through links on a map of the SRER. All UTM coordinates for these measurement locations are presented in text format (.txt). The annotated bibliography is hyperlinked text, and updates are .pdf files. Supporting spatial data sets are available as ArcInfo export format (.e00) and the associated metadata in text (.txt) format. Metadata for the vegetation abundance include the descriptions of data source, history, measurement units, missing values, and attribute definitions and value descriptions in file rows and columns. Metadata for the spatial coverages include the date of file construction, characteristics of the projection, data type, description of content, file format, file size, history of file development and sources of data, and attribute definitions and value descriptions in file rows and columns. Currently, the metadata do not meet any quality standards, but the spatial coverage metadata are being updated to meet the FGDC standards.

The data are stored on the CALS servers (http://cals.arizona.edu/srer). The Director for Research at SRER maintains the site with assistance from the CALS Network Laboratory, but there is no budget dedicated for archive maintenance.

Approximately \$260 000 has been spent developing, updating, and maintaining the digital archive. Conversion to digital files, including data entry into spreadsheet and text files, was supported by \$60 000 from the UA Vice President for Research in 1994. Repeat photography scanning was donated by the USDA-ARS Jornada Experimental Range in 1996. The USDA Forest Service Rocky Mountain Research Station has provided ~\$16 000 annually since 1995 to develop digital content, remeasure vegetation transects (every three years), and rephotography at photo stations (twice since 1995). The original Web site design was provided without charge by the UA Office of Arid Lands Studies in 1996.

Future funding is uncertain because resources from USDA Forest Service Rocky Mountain Research Station are based on the competitive reviews and availability of funding. Nevertheless, returns on these funds include a revitalization of ongoing research and the recent selection as a Core Site for the Southwest Desert Domain in the emerging National Ecological Observatory Network. Prior to 1994, there were less than ten ongoing research projects and very little recognition of research opportunities beyond the local community. Since 2000, there has been, on average, 35 ongoing projects, with about a six-to-eight project turnover annually, and over 100 peer-reviewed publications from works performed on the SRER and/or using the digital archive. Scientists from other states and countries now use the SRER as well. The data are also made available through the Knowledge Network for Biocomplexity, which is a clearinghouse of "ecological data from a highly distributed network of field stations, laboratories, research sites, and individual researchers" (http://knb.ecoinformatics.org).

C. SWRC DAP and WIGDA

The SWRC DAP is an electronic data processing system that includes an online interface (http://tucson.ars.ag.gov/dap) to provide public access to the data collected at WGEW and SRER by SWRC [42]. The DAP project currently supports access to long-term measurements of precipitation, runoff, sediment, meteorological conditions, soil moisture and temperature, vegetation, CO_2 flux, and evapotranspiration [38].

The SWRC DAP was initiated in 2000 by a team of SWRC scientists, information technology (IT) specialists, hydrological technicians, and instrumentation specialists, with the goal to "make high-quality data available to users for scientific research." The DAP team identified over 20 SWRC data sets that had potential for archiving and distribution but quickly came to agreement that rainfall and runoff, sediment, and meteorological data had the highest priority. This was based on a formal assessment of several criteria including relevance to the SWRC mission, length of continuous data record, potential for future measurements, data quality, percentage of data already in digital format, and status of accessibility. Of these three data sets, the rainfall and runoff data set was selected for DAP with the simple achievable goal to have all WGEW and SRER rainfall and runoff data archived in digital format and accessible online by October 2003—the fiftieth anniversary of WGEW and the one hundredth anniversary of SRER.

A significant effort of SWRC scientists and IT specialists over 1.5 years went into designing the overall data table structure and logic, including table definitions and relationships. During this stage, decisions were made about data processing consistency (i.e., historic data would not be reprocessed and only gross errors would be corrected), quality checking (i.e., it would be largely automated, requiring minimal human attention), metadata format to be followed (i.e., data descriptions would be published in a peer-reviewed journal), etc. A more intensive effort over 2.5 years was invested in the implementation of the DAP plan. During this stage, the DAP team of scientists, technicians, IT specialists, and management met weekly to allow work (e.g., database design, data archive, and data quality checking) to be done in parallel with coordination and communication. DAP became an integral part of unit decisions about hiring and supervision, and garnered a realistic budget. The project was completed nearly on schedule, in May 2004. At that point, a DAP follow-on plan was initiated to add six new databases, including GIS data and measurements of sediment, soil moisture, meteorological data, energy flux, and vegetation, with a set deadline of September 2007. For this DAP follow-on project, some temporary staffs were added, and the project was completed by January 2008. The total investment in DAP to put eight data sets up on the web with metadata published in a peer-reviewed journal [38] was \$690 800 (Table I; publication costs not included). DAP maintenance requires some leadership, IT upgrades, limited data quality checking, and maintenance of data loggers and instrumentation at a cost estimated to be \$20 000/year.

The DAP webserver is a Dell Server running MS Windows Server 2000 operating system that is located outside of the internal SWRC network. The interactive web interface for the DAP was developed using Visual Basic Script and MS Active Server Pages. MS Active Server Pages is the webserver scripting environment that is a part of the MS Internet Information Services webserver software that is included with MS Windows Server products. The graphical access to the data was developed

TABLE I
BREAKDOWN OF INVESTMENT IN SWRC DAP AND WIGDA

Task	Task Duration	SWRC Team	Temporary Staff	Task Cost	Task Product
DAP Planning	1.5 years	3 Scientists, 1 Support Scientist		$92,000	1) DAP 3-year plan 2) DAP database and interface design
DAP Implementation	2.5 years	3 Scientists, 2 Support Scientists, 2 IT Specialists, 1 Technician		$324,800	1) DAP completed 2) 50 years of rainfall and runoff available on web
DAP Follow-on	3.5 years	1 Scientists, 1 Support Scientist, 2 Technicians, 1 IT Specialist	1 IT Specialist, 3 students	$274,000	1) Map interface added to DAP 2) Six new datasets added to DAP – sediment, soil moisture, met data, GIS, energy flux and vegetation
DAP Maintenance	Ongoing, starting in 2008	1 Scientist, 1 IT Specialist, 1 Technician		$20,000 per year	1) Updates to data and system technology
WIGDA Implementation	1.0 year	1 Scientist, 1 Technician	1 IT Specialist	$111,100	1) a relational database of image and ground measurement metadata
WIGDA Maintenance	Ongoing, starting in 2008	1 Technician		$1,000 per year	1) up-to-date metadata on SWRC geospatial data holdings

using ESRI ArcIMS. DAP data are processed into a Microsoft Structured Query Language (SQL) Server and stored on a file server with RAID redundancy (a detailed architectural diagram of DAP is given by Nichols and Anson [42]). The files that make up the SQL Server database are made public and backed up to the main file server nightly. Symantec Backup Exec software is used to perform daily incremental backups of the main file server onto tape (LTO-3). Each month, the entire database is written to a set of tapes and rotated to an offsite location.

The publicly accessible part of SWRC DAP consists of an interactive Web site, which provides an interface to the data and metadata, and a relational database to process, store, and manage data. The graphical interface integrates high resolution orthophotos, Landsat satellite images, and GIS data layers containing watershed boundaries, channel networks, and point locations of flumes and rain gauges. After the user selects instrumented sites through the graphical display, the site parameters are used to populate the standard form, and data are displayed in response to submission (Fig. 4). The DAP system was released to the public in October 2003, and since that time, the online data access Web site has received nearly 5000 visitors.

Coincident with DAP was the SWRC development of the WIGDA. WIGDA is an archive of metadata that can be used to locate comprehensive sets of image and ground data files associated with remote sensing field campaigns at WGEW and the surrounding San Pedro River Basin over the past 20 years. The metadata schema established in 1997 was patterned after the ARIA metadata schema for searching and locating images matching user-specified requirements, as described earlier in the ARIA section. WIGDA includes metadata for hundreds of spectral images from a variety of satellite- and aircraft-based sensors (the "images") associated with high-quality ground-based measurements of soil, plant, and atmospheric conditions

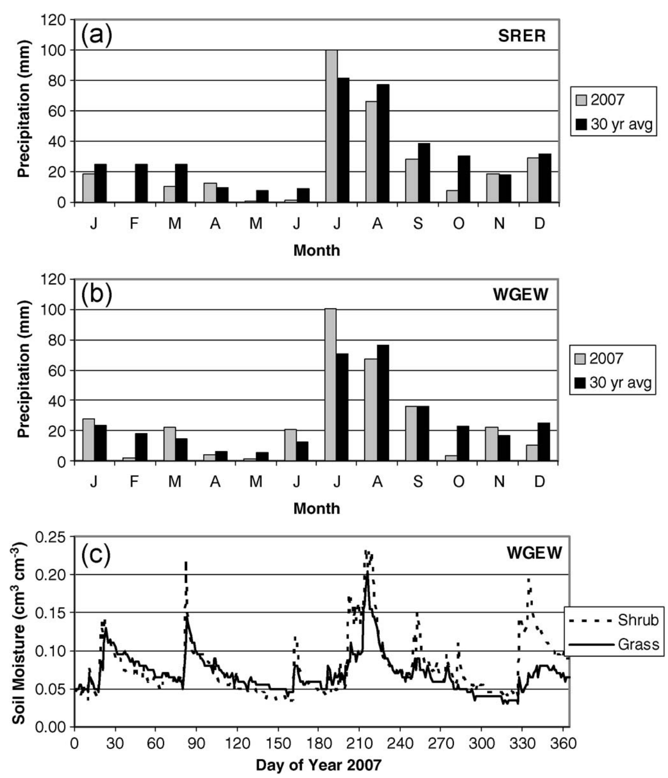

Fig. 4. Examples of data available through SWRC DAP. Average monthly precipitation data compiled for multiple rain gauges in (a) SRER and (b) WGEW, for the 30-year time period encompassing most ARIA image acquisitions and the most recent year 2007, and (c) the volumetric soil moisture to a depth of approximately 15 cm, measured daily in WGEW in 2007 at shrub- and grass-dominated sites, illustrating the spatial variability induced by localized storms across the watershed.

(the "ground data"). Users can select a particular image file and see metadata for all related ground data files (and *vice versa*) to locate data sets that meet specific research needs.

This database of metadata was published under the name "WIGDA 1999" or WIGDA99 [41]) and later replaced with WIGDA06 [39]. The WIGDA06 interface is programmed in Visual Basic.Net for Microsoft.NET Framework 2.0 and available for download from the DAP Web site at http://www.tucson.ars.ag.gov/dap/. The total investment in WIGDA06 (and its predecessor WIGDA99) was estimated at $111 100, i.e., the salary of a high-level technician for one year and an experienced programmer for one half year. The annual maintenance is only the addition of metadata for new images and ground-based data collection, at a cost of about $1000/year (Table I).

Interconnectivity with other data distribution systems was key to DAP success. For example, the WIGDA06 database provided the metadata link between images and image-related ground data but did not have the capacity for interactive image downloading. To resolve this, the images identified by WIGDA06 were made available from ARIA, and the WIGDA06 database was revised to include a column identifying those images available for download from ARIA. Then, the SWRC Web site was linked to the ARIA Web site to allow easy distribution of the selected images, and the ARIA Web site was revised to allow the WIGDA06 database to be downloaded directly from ARIA.

Rather than support redundant data holdings, the DAP Web site provides the GIS layers associated with ARIA spectral images, and conversely, the ARIA Web site provides the DEM and DOQQ images associated with SRER and WGEW long-term measurements. The long-term data archived and distributed by SRER DD and SWRC DAP offer the opportunity for quantitative analysis of ARIA images (e.g., [28]). Links between ARIA, SRER, and WGEW Web sites facilitate data distribution for scientists unfamiliar with the area.

The immediate return on DAP and WIGDA investment has been substantial. As a direct result of DAP, SWRC was able to compare measurements from a redundant set of digital and analog rain gauges [31] and determine that the measurement of rainfall with analog rain gauges could be discontinued, resulting in a total savings to the unit of over $100 000/year. Thus, based only on this tangible savings, the cost of DAP would be recuperated in only seven years. Another direct result of DAP was a striking increase in SWRC productivity; in 2008, 18 manuscripts authored by SWRC scientists will be published in a Special Section of the journal *Water Resources Research* dedicated to WGEW long-term data and the DAP Web site. Upon completion of WIGDA99, a proposal was funded by NASA to use the images and ground-based data organized in WIGDA99 and available in ARIA for "EO-1 and Landsat intersatellite comparison at two established Arizona field sites" [4]. Thus, a unit investment of $111100 was recognized by a funding agency to be worth a soft-money investment of over $300 000 for research.

The DAP was designed for sustainability through the foreseeable future. That is, the data updating scheme is largely automated and can be accomplished with a small time investment from permanent staff. It is envisioned that permanent staff will continue to upgrade media, hardware, and software as technology improves with time. The DAP Web site was added to the American Geophysical Union-approved data sites based on the following criteria: "1) access to data should be unrestricted and at reasonable or no cost and 2) the data centers should be committed to the permanent maintenance of data sets within their mandate" (http://www.agu.org/pubs/datacent.html).

V. Discussion

The approaches of government and academic units in southeastern Arizona for archiving and distributing long-term geospatial data had several elements in common. First, all recognized the importance of providing digital archives with user-friendly interfaces for the various data sets. In addition, they were based on clear and realistic goals with contingencies for scaling back or ramping up, depending on budget and interest. Each effort began with a strong advocate(s), which attracted inter- and intraunit support for the project and provided coordination and leadership. Data archiving and distribution were an interdisciplinary effort involving knowledge of science, instrumentation, research, IT, history, and personnel. Permanent staff held the decision-making responsibility, and temporary hires were used for well-defined tasks.

However, there was a distinction between approaches by government and academic units to data archiving and distribution, defined by inherent institutional differences. Efforts by the academic units were initiated by short-term "seed" money that was not premised on plans for long-term sustainability. Conversely, efforts by the government unit were funded by diverting existing unit funds to the effort, with a plan to sustain the system with permanent staff dedicated to the unit mission. This is an important distinction that helps explain the basic difficulty and inaction in long-term geospatial data preservation in both government and academic units. The government unit had trouble initiating the project but had the capacity to sustain it indefinitely; the academic units had means to initiate projects but the prospects for long-term sustainability and upgrade of the systems were precarious. Despite the differences in approaches for long-term data archiving and distribution, the results were the same for government and academic units. Robust systems were developed for distributing diverse long-term geospatial data with good prospects for long-term sustainability. On the other hand, these localized efforts resulted in a general lack of attention to standardizing preservation metadata formats such as the OAIS model.

Approaches to sustaining these systems differed between government and academic units. At the government unit, sustaining and upgrading the system are a unit-wide effort supported with unit funds; at the academic units, sustainability is an interunit effort based on ongoing solicitation of support from local federal units, other academic units, and users. In both cases, sustainability relies upon a widespread recognition that continued access to the long-term data is valuable. Each of the example cases had quantifiable impact on publications, budget savings, outside recognition, student research, registered users, data downloads, and new research. The value is also determined by the potential for data collection, archiving, and distribution to be continued, leading to the positive feedback of sustainability feeding sustainability. That is, the longer a data archiving and distribution system has been in place, the more interest and investment it attracts to sustain it. High impact also serves to attract more advocates, which can increase the chances of long-term sustainability despite changes in personnel and technology. Nevertheless, sustainability in the academic setting remains problematic when university budgets and discretionary funds are being cut at the same time that federal and state research funding levels fall. Such is the case now in Arizona, adding increased uncertainty about the future of these resources.

At the same time, success in long-term data archiving and distribution does not guarantee the interconnection of the geospatial data. Interconnection is the process of linking the data distribution systems across platforms to allow the search and retrieval of interdisciplinary data sets for new research that would not be possible with one data set alone. Many approaches for interconnection were explored in the case studies given here. For example, the Web sites for ARIA, SRER DD, and WGEW DAP have Web site links, one to another, which required only a 1-h effort by each webmaster. This allows users to access the SRER and WGEW GIS layers that are associated with ARIA image holdings and the ARIA DOQQ images that can be used to interpret SRER and WGEW tabular measurements. The satellite images acquired and archived by SWRC are distributed by ARIA based on an agreement that SWRC would fund a technician to upload SWRC images to ARIA. This combined the ongoing acquisition and archiving of images by SWRC and the existing ARIA image distribution system to leverage the strength of both. At a greater investment (Table I), the WIGDA relational database was developed to link image and ground metadata to allow the compilation of interdisciplinary data sets that meet specific research needs. Interconnection was also achieved at little cost by submitting subsets of long-term data archived by the local units to national repositories of publicly accessible data sets. For example, the SRER and WGEW databases of hydrological and ecological measurements are available through the previously mentioned EcoTrends project (http://www.ecotrends.info), which offers access to a large collection of long-term geospatial data from dozens of sites across the U.S. The advantage of repositories like EcoTrends is that long-term data from a multitude of sites can be downloaded in a uniform format, ready for analysis, with the same units, and temporal frequency.

The interconnection of the ARIA, SRER DD, and SWRC DAP data distribution systems has already produced results. A ten-year record of NASA Landsat images accessible through ARIA was combined with long-term meteorological and energy flux measurements available from SWRC DAP to map the spatial and temporal distribution of net daily CO_2 flux at WGEW [62]. NASA Landsat images (1989, 1994, and 2005) from ARIA were combined with DEM data from SRER DD to support an assessment of errors associated with the estimates of carbon mass woody plants on the SRER following wildfires [28]. In addition, the 50-year record of precipitation at SRER and WGEW was downloaded from the EcoTrends project, along with similar records from 11 other sites, to explore the variability in mean annual precipitation across the conterminous U.S. [40].

VI. Conclusion

The results from these three cases demonstrate that the challenges of digitizing, archiving, connecting, and providing access to geospatial data can be successfully met at the local level, although some challenges remain. With the concerted and combined efforts of many advocates, the problems of and opportunities for making these data sets available for current and future generations of researchers and students were recognized and at least partially addressed. Although it is clear that there are advantages for government agencies in the long term, as more stable budgets can be leveraged for archival activities, the growing understanding of the issue in the academic community, particularly among libraries and research centers, along with technical advances for creating greater interoperability and new consortial responses for data storage, bode well for more support and options in the coming years. However, to ensure this progress, there must be continued policy development at the national level and a determination from agencies funding scientific research that costs for data archiving and access are included in their programs.

Acknowledgment

The SWRC DAP is a product of contributions of the *entire* SWRC staff, past and present, who made it their mission to maintain quality long-term measurements and provide access to these data online for the common good. The authors would like to thank S. C. Martin, C. Hutchinson, R. MacArthur, C. Edminster, and E. Pfirman for providing the initial nurturing needed to support the DD for the SRER, and also E. Pfirman, who was the original architect of ARIA.

References

[1] J. D. All, “Colorado river floods, droughts, and shrimp fishing in the upper gulf of California, Mexico,” *Environ. Manage.*, vol. 37, no. 1, pp. 111–125, Jan. 2006.

[2] C. H. Atkinson and C. E. Isbell, “Preservation of digital data from high volume data repositories,” in *Proc. Amer. Geophys. Union Conf.*, San Francisco, CA, 2003. [Online]. Available: http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/8024/1/03-3319.pdf

[3] D. R. Bleakly, “Long-term spatial data preservation and archiving: What are the issues?” Sandia Nat. Lab., Albuquerque, NM, SAND Report 2002–0107, 2002.

[4] R. Bryant, M. S. Moran, S. A. McElroy, C. Holifield, K. J. Thome, T. Miura, and S. F. Biggar, “Data continuity of Earth Observing 1 (EO-1) Advanced Land I satellite imager (ALI) and Landsat TM and ETM+,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1204–1214, Jun. 2003.

[5] Cornell Chronicle Online, *$400000 NSF grant will aid sharing of raw research data*, Nov. 20, 2007.

[6] Council on Library and Information Resources and the Library of Congress, *Building a national strategy for digital preservation: Issues in digital media archiving*, 2002, Washington DC: Council on Library and Information Resources. [Online]. Available: http://www.clir.org/pubs/reports/pub106/contents.html

[7] R. Crow, “*The Case for Institutional Repositories: A SPARC Position Paper*. Washington DC: The Scholarly Publishing & Academic Resources Coalition, 2002. [Online]. Available: http://works.bepress.com/ir_research/7/

[8] M. Datcu, S. D’Elia, R. L. King, and L. Bruzzone, “Introduction to the special section on image information mining for Earth observation data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 795–798, Apr. 2007.

[9] R. Dekker, “The importance of having data-sets,” in *Proc. Conf. Int. Assoc. Technol. Univ. Libr.*, 2006, pp. 1–4. [Online]. Available: http://www.iatul.org/doclibrary/public/Conf_Proceedings/2006/Dekkerpaper.pdf

[10] M. de Martino, S. B. Serpico, and C. Camurri, “Forest mapping by partially supervised classification applied to vegetation indexes,” in *Proc. IGARSS*, Jul. 25–29, 2005, vol. 5, pp. 3128–3131.

[11] J. N. DiBari, “Evaluation of five landscape-level metrics for measuring the effects of urbanization on landscape structure: The case of Tucson, Arizona, USA,” *Landsc. Urban Plan.*, vol. 79, no. 3/4, pp. 308–313, 2006.

[12] J. E. Diem and A. C. Comrie, “Integrating remote sensing and local vegetation information for a high-resolution biogenic emissions inventory—application to an urbanized, semiarid region,” *J. Air Waste Manage. Assoc.*, vol. 50, no. 11, pp. 1968–1979, Nov. 2000.

[13] J. E. Diem, “Remote assessment of forest health in Southern Arizona, USA: Evidence for ozone-induced foliar injury,” *Environ. Manage.*, vol. 29, no. 3, pp. 373–384, Mar. 2002.

[14] J. E. Diem and A. C. Comrie, “Predictive mapping of air pollution involving sparse spatial observations,” *Environ. Pollut.*, vol. 119, no. 1, pp. 99–117, 2002.

[15] L. Diranti, “The long-term preservation of accurate and authentic digital date: The InterPARES project,” *Data Sci. J.*, vol. 4, pp. 106–118, 2005.

[16] P. Doorn and H. Tjalsma, “Introduction: Archiving research data,” *Arch. Sci.*, vol. 7, no. 1, pp. 1–20, Mar. 2007.

[17] *EcoTrends project website*. [Online]. Available: http://www.ecotrends.info/EcoTrends/index.jsp

[18] Federal Geographic Data Committee, *Content standard for digital geospatial metadata—FGDC geographic metadata committee*, 1998. [Online]. Available: http://www.fgdc.gov/metadata/csdgm/

[19] Federal Geographic Data Committee, *FGDC/ISO metadata standard harmonization website*, 2006. Accessed on June 1, 2008. [Online]. Available: http://www.fgdc.gov/metadata/us-national-profile-iso19115/archive

[20] S. E. Flores and S. R. Yool, “Sensitivity of change vector analysis to land cover change in an arid ecosystem,” *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 1069–1088, Jan. 2007.

[21] E. P. Glenn, P. L. Nagler, R. C. Brusca, and O. Hinojosa-Huerta, “Coastal wetlands of the northern Gulf of California: Inventory and conservation status,” *Aquat. Conserv.*, vol. 16, no. 1, pp. 5–28, 2006.

[22] E. P. Glenn, K. W. Flessa, M. J. Cohen, P. L. Nagler, K. Rowell, and F. Zamora-Arroyo, “Just add water and the Colorado river still reaches the sea,” *Environ. Manage.*, vol. 40, no. 1, pp. 1–6, Jul. 2007.

[23] D. C. Goodrich, T. O. Keefer, C. L. Unkrich, M. H. Nichols, H. B. Osborn, J. J. Stone, and J. R. Smith, “Long-term precipitation database, Walnut Gulch Experimental Watershed, Arizona, United States,” *Water Resour. Res.*, vol. 44, p. W05 S04, May 2008. DOI: 10.1029/2006WR005782.

[24] S. Hettich and S. D. Bay, *The UCI KDD Archive*. Irvine, CA: Dept. Inf. Comput. Sci., Univ. California, 1999. [Online]. Available: http://kdd.ics.uci.edu

[25] P. Heilman, M. H. Nichols, D. C. Goodrich, S. N. Miller, and D. P. Guertin, “Geographic information systems database, Walnut Gulch Experimental Watershed, Arizona, United States,” *Water Resour. Res.*, vol. 44, p. W05 S11, 2008. DOI: 10.1029/2006WR005777.

[26] M. Herald, C. E. Woodcok, A. di Gregorio, P. Mayaux, A. S. Belward, J. Latham, and C. C. Schmullius, “A joint initiative for harmonization and validation of land cover datasets,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1719–1727, Jul. 2006. DOI:10.1109/TGRS.2006.871219.

[27] C. Huang, L. Geiger, and J. A. Kupfer, “Sensitivity of landscape metrics to classification scheme,” *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 2927–2948, 2006.

[28] C. Y. Huang, S. E. Marsh, M. P. McClaran, and S. R. Archer, “Postfire stand structure in a semiarid savanna: Cross-scale challenges estimating biomass,” *Ecol. Appl.*, vol. 17, no. 7, pp. 1899–1910, Oct. 2007.

[29] C. F. Hutchinson, S. E. Marsh, B. Orr, P. Krausman, R. M. Enns, L. Howery, E. Pfirman, C. Wallace, J. Walker, K. Mauz, H. Boyd, H. Salazar, and E. Trobia, “Informing the elk debate: Applying NASA Earth Observing System (EOS) data to natural resource management conflicts in the western states,” in *Proc. IGARSS*, 2001, vol. 2, pp. 834–836.

[30] T. O. Keefer, M. S. Moran, and G. B. Paige, “Long-term meteorological and soil hydrology database, Walnut Gulch Experimental Watershed, Arizona, United States,” *Water Resour. Res.*, vol. 44, p. W05 S07, 2008. DOI:10.1029/2006WR005702.

[31] T. O. Keefer, C. L. Unkrich, J. R. Smith, D. C. Goodrich, M. S. Moran, and J. R. Simanton, “An event-based comparison of two types of automated-recording, weighing bucket rain gauges,” *Water Resour. Res.*, vol. 44, no. 5, p. W05 S12, May 2008. DOI:10.1029/2006WR005841.

[32] G. King, “An introduction of the Dataverse Network as an infrastructure for data sharing,” *Sociol. Methods Res.*, vol. 36, no. 2, pp. 173–199, Nov. 2007.

[33] B. Lavoie, “Meeting the challenges of digital preservation: The OAIS reference model,” *OCLC Newsl.*, vol. 243, pp. 26–30, Jan./Feb. 2000.

[34] C. A. Lynch, “Institutional repositories: Essential infrastructure for scholarship in the digital age,” *Libr. Acad.*, vol. 3, no. 2, pp. 327–336, Apr. 2003.

[35] A. Madrigal, “Google to host terabytes of open-source science data,” *Wired Science–Wired Blogs*, 2008. [Online]. Available: http://blog.wired.com/wiredscience/dataset/index.html

[36] M. P. McClaran, D. L. Angell, and C. Wissler, “Santa Rita experimental range digital database: User’s guide,” USDA, Forest Service, Rocky Mountain Res. Station, Ogden, UT, General Tech. Rep. RMRS-GTR-100, 2002.

[37] A. L. Medina, “The Santa Rita experimental range: History and annotated bibliography (1903–1988),” USDA, Forest Service, Rocky Mountain Res. Station, Ogden, UT, General Tech. Rep. RM-GTR-276, 1996.

[38] M. S. Moran, W. E. Emmerich, D. C. Goodrich, P. Heilman, C. D. 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. Wong, “Preface to special section on fifty years of research and data collection: U.S. Department of Agriculture Walnut Gulch Experimental Watershed,” *Water Resour. Res.*, vol. 44, p. W05 S01, 2008. DOI:10.1029/2007WR006083.

[39] M. S. Moran, C. D. Holifield Collins, D. C. Goodrich, J. Qi, D. T. Shannon, and A. Olsson, “Long-term remote sensing database, Walnut Gulch Experimental Watershed, Arizona, United States,” *Water Resour. Res.*, vol. 44, p. W05 S10, 2008. DOI:10.1029/2006WR005689.

[40] M. S. Moran, D. C. P. Peters, M. P. McClaran, M. H. Nichols, and M. B. Adams, “Long-term data collection at USDA experimental sites for studies of ecohydrology,” *J. Ecohydrol.*, 2008. in press.

[41] M. S. Moran, J. Qi, W. Ni, and D. T. Shannon, “Water conservation laboratory image and ground data archive (WIGDA99) for Arizona agricultural and rangeland regions,” in *Proc. 2nd Int. Conf. Geospatial Inf. Agric. For. 10–12 Jan.*, Orlando, FL, 2000, pp. II:229–II:236.

[42] M. H. Nichols and E. Anson, “Southwest Watershed Research Center data access project,” *Water Resour. Res.*, vol. 44, p. W05 S03, 2008. DOI:10.1029/2006WR005665.

[43] S. P. Morris and J. Tuttle, “Curation and preservation of complex data: The North Carolina geospatial data archiving project,” in *Proc. Digital Content Curation Symp.*, 2007. [Online]. Available: http://repository.lib.ncsu.edu/publications/bistream/1840.2/105/1/digccurr2007ncgdap.pdf

[44] L. M. Namikawa and C. S. Renschler, “Dynamic integrated GIS enhancement and support tools,” in *Proc. GEOINFO*, Saõ Paulo, Brazil, Nov. 3–5, 2003.

[45] National Research Council (US), *Issues for science and engineering researchers in the digital age: Office of special projects*, 2001, Washington DC: National Academy Press.

[46] National Research Council (US), *Committee on the preservation of geoscience data and collections. Geoscience data and collections: National resources in peril*, 2002, Washington DC: National Academy Press.

[47] “Committee on archiving and accessing environmental and geospatial data at NOAA, board on atmospheric sciences and climate, division on earth and life studies,” *Environmental Data Management at NOAA: Archiving, Stewardship, and Access*, 2007, Washington DC: National Academy Press.

[48] National Science Foundation (US), “Sustainable digital data preservation and access network partners (DataNet), office of cyberinfrastructure,” *Program Synopsis*, 2008. [Online]. Available: http://www.nsf.gov/funding/pgm_sum.jsp?pims_id=503141

[49] L. N. Norman, A. J. Donelson, E. L. Pfeifer, A. H. Lam, and K. J. Osborn, “Monitoring colonias development along the United States–Mexico border: A process application using GIS and remote sensing in Douglas, Arizona, and Agua Prieta, Sonora,” USGS, Menlo Park, USGS Open-File Report 2004-1212, May 2004.

[50] OCLC/RLG Working Group on Preservation Metadata, *Preservation metadata and the OAIS information model: A metadata framework to support the preservation of digital objects*, 2002. [Online]. Available: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

[51] F. Olsen, *Digital Archiving: Ensuring Storage Space and Access*. Washington DC: Chronicle Higher Educ., Jan. 30, 2004, p. 14.

[52] O. Y. Rieger, *Preservation in the age of large-scale digitization: A white paper*, 2007. Council Library Inf. Res. [Online]. Available: http://www.clir.org/pubs/abstract/pub141abst.html

[53] K. G. Renard, M. H. Nichols, D. A. Woolhiser, and H. B. Osborn, “A brief background on the U.S. Department of Agriculture Agricultural Research Service Walnut Gulch Experimental Watershed,” *Water Resour. Res.*, vol. 44, p. W05 S02, 2008. DOI:10.1029/2006WR005691.

[54] J. Robertson, *Geoscience data preservation. GSA position statement*, 2005. The Geological Soc. Amer. [Online]. Available: http://www.geosociety.org/positions/pos9_dataPres.pdf
[55] J. Rose, "Preservation of original digital field-recorded time-series data," *Water Resources Division Memorandum No. 99.33*, Sep. 27, 1999. [Online]. Available: http://water.usgs.gov/admin/memo/policy/wrdpolicy99.33.html
[56] B. C. Rundquist, C. J. Henrie, and E. J. Grewe, "Internet access to remotely sensed data satellite imaging made commonplace," *J. Map Geogr. Libr.*, vol. 2, pp. 21–32, 2006.
[57] "Digital preservation: Alliance set to tackle science's new frontier," *Science Daily*, Dec. 7, 2007. [Online]. Available: http://www.sciencedaily.com/releases/2007/12/071204180015.htm
[58] J. J. Stone, M. H. Nichols, D. C. Goodrich, and J. Buono, "Long-term runoff database, Walnut Gulch Experimental Watershed, Arizona, United States," *Water Resour. Res.*, vol. 44, p. W05 S05, 2008. DOI:10.1029/2006WR005733.
[59] University of Michigan, *Deep blue institutional repository: Best practices for producing datasets*, 2006. [Online]. Available: http://hdl.handle.net/2027.42/40246
[60] *Using hydrogeomorphic surfaces for delineating floodplains: Black water creek test reach within the upper Puerco Watershed, Navajo Nation*, 2004. Hanover, NH: US Army Corps Eng. Cold Regions Res. Eng. Lab. [Online]. Available: ERDC/CRREL TN-04-7.
[61] A. B. White and P. Kumar, "A data mining approach for understanding topographic control on climate-induced inter-annual vegetation variability over the United States," *Remote Sens. Environ.*, vol. 98, no. 1, pp. 1–20, Sep. 2005.
[62] C. Holifield Collins, W. E. Emmerich, M. S. Moran, M. Hernandez, R. L. Scott, R. Bryant, D. M. King, and C. L. Verdugo, "A remote sensing approach for estimating distributed daily net carbon dioxide flux in semiarid grasslands," *Water Resour. Res.*, vol. 44, p. W05 S17, 2008. DOI: 10.1029/2006WR005699.

Mary Susan Moran received the Ph.D. degree in soil and water science from The University of Arizona, Tucson, in 1990.

She is currently a Hydrologist with the Southwest Watershed Research Center, Agricultural Research Service, U.S. Department of Agriculture, Tucson, Arizona. She also holds an adjunct faculty appointment with the Department of Soil, Water, and Environmental Science, The University of Arizona. Her research is concentrated on critical arid and semiarid regions with a broad focus on the impact of global change on natural resources management and a specialized focus on the application of remote sensing to difficult problems. She has also served on the NASA Landsat Science Team and EO-1 Validation Team to evaluate selected technologies for meeting soil science needs in the twenty-first century.

Barbara S. Hutchinson received the B.A. degree in history from the University of California, Riverside, the M.S. degree in library science, and the Ph.D. degree in higher education, from The University of Arizona, Tucson. Her doctoral studies in higher education were focused on organizational change and the role of academic research units.

She is currently an Assistant to the Dean and Director of the Agricultural Experiment Station, College of Agriculture and Life Sciences (CALS), The University of Arizona. In that capacity, she provides grant support to CALS faculty and develops international extramural programs. The areas of research interest include the electronic dissemination of information and educational resources, technologies for facilitating life-long learning and decision making, and digital repositories. These interests have been most fully realized in her involvement in the Agriculture Network Information Center (AgNIC) and the development of the Rangelands West initiative that involves 19 land-grant universities in a model collaboration of librarians and rangeland scientists. She is also a Professor in agricultural education.

Dr. Hutchinson has served as the President of the U.S. Agricultural Information Network and the Chair of the AgNIC Executive Board and is now a Board member of the International Association of Agricultural Information Specialists and a member of the North American Advisory Board for CAB International. She has also provided technical assistance in information management in countries throughout the world including Bahrain, Botswana, Malawi, Kenya, India, and Chile.

Stuart E. Marsh received the B.S. degree in geology from George Washington University, Washington, DC, in 1973 and the M.S. and Ph.D. degrees in applied earth sciences from Stanford University, Stanford, CA, in 1975 and 1979, respectively.

He is currently a Professor and the Chair of the Arid Lands Resource Sciences Interdisciplinary Ph.D. Program, and the Director of the Arizona Remote Sensing Center, Office of Arid Lands Studies, The University of Arizona, Tucson. He also holds a joint faculty appointment with the Department of Geography and Regional Development. He is currently involved in a broad range of research designed to assess surface environmental conditions utilizing advanced remote sensing and geographic information system technologies. These research efforts have attempted to enhance our abilities to employ these technologies to address the impacts of land use and land cover change and environmental degradation. Recent published research has focused on the following: 1) the discrimination of invaded and native species sites using multitemporal remote sensing; 2) understanding anthropogenic impacts on estimating biomass in semiarid environments; 3) estimating the impacts of grazing and fuel–wood management on soil–carbon sequestration; 4) biophysical characterization and management effects on semiarid rangelands using Landsat satellite data; and 5) characterizing the spatial structure of endangered species habitat using geostatistical analysis.

Mitchel P. McClaran received the Ph.D. degree from the University of California, Berkeley, in 1986.

Since 1986, he has been a member of the faculty with The University of Arizona (UA), Tucson, where he is currently a Professor of range management and the Chair of the Graduate Program in Natural Resource Studies, School of Natural Resources, and the Director for Research at the 21 500-ha UA Santa Rita Experimental Range. He teaches and researches rangeland ecology and policy.

Dr. McClaran was the recipient of the Outstanding Teaching Award from the Range Science and Education Council in 1999, the School of Natural Resources Outstanding Scholarship Achievement in 2004, and the Outstanding Academic Book (The Desert Grassland) by Choice magazine in 1997.

Aaryn D. Olsson received the B.S. degree in mathematics and the M.S. degree in renewable natural resource studies from The University of Arizona, Tucson, in 2001 and 2006, respectively, where he is currently working toward the Ph.D. degree in the Arid Lands Resource Sciences Program.

He is also a Senior Research Specialist with the Arizona Remote Sensing Center, Office of Arid Lands Studies, The University of Arizona, currently researching ecological and societal impacts of invasive species with a focus on remote sensing, geographic information systems, decision support systems, and dynamic programming. Before receiving the M.S. degree, he managed the Arizona Regional Image Archive as both System Manager and Programmer, in addition to serving as Lead Programmer on a number of online spatial decision support systems.