System Requirements Specification

for

STEWARDS

Version 1.0 approved

Prepared by Conservation Effects Assessment Project (CEAP)

Data System Team

USDA-Agricultural Research Service

Table of Contents

Table of Contents ..... **ii**
Revision History ..... **iii**
1. Introduction **1**
1.1 Purpose 1
1.2 About this Document and its Readers 1
1.3 Project Scope 1
1.4 References 2
2. Overall Description **3**
2.1 Product Perspective 3
2.2 Product Features 3
2.3 User Classes and Characteristics 4
2.4 Operating Environment 4
2.5 Design and Implementation Constraints 5
2.6 User Documentation 5
2.7 Assumptions and Dependencies 6
3. System Features **6**
3.1 Database Management System 6
3.1.1 Description and Priority 6
3.1.2 Use Case Scenarios 6
3.1.3 Functional Requirements 6
3.2 Metadata Availability 8
3.3 Data Dissemination 9
3.4 Hardware requirements 9
3.5 Software requirements 9
3.6 Data migration to and retrieval from STEWARDS 10
3.6.1 Quality Assurance 10
3.6.2 Populating data 10
3.6.3 Updating Data 11
3.6.4 Data derivation/aggregation 11
4. External Interface Requirements **11**
4.1 User Interfaces 11
4.2 Hardware Interfaces 11
4.3 Software Interfaces 12
4.4 Communications Interfaces 12
5. Other Nonfunctional Requirements **12**
5.1 Performance Requirements 12
5.2 Safety Requirements 12
5.3 Security Requirements 12
5.4 Software Quality Attributes 13
Appendix A: Acronyms and Glossary **13**
Appendix B: Analysis Models **16**
Appendix C: Issues List **16**
Appendix D: Database team members **16**
Appendix E: System Detailed Specifications **17**
Appendix F: System Design **18**
Appendix G: Linkage of STEWARDS to Other USDA/ARS Data/Modeling Systems **39**
Appendix H: Task Flow of STEWARDS **40**
Appendix I: Data Structure for Two Types of Time Series Data in STEWARDS **41**

Revision History

Name	Date	Reason For Changes	Version
Sadler	11/10/04	First draft of assigned sections	0.1
Sadler	12/08/04	Third draft of assigned sections	0.3
James	2/8/05	Contributions	0.3b
CEAP team	3/3/05	Discussions at Irving CEAP meeting	d0.5
Sadler	3/8/05	Continued from Irving and other observations	0.6
James	3/10/05	Ongoing edits	0.61
Sadler	3/11/05	Added Jin-Song's sections and Pruitt's comments, accepted some of Dave's to make it easier to see the material changes.	0.62
OCIO	3/22/05	OCIO inputs	0.63
Sadler	3/29/05	Address upload responsibilities, MP comments except those OCIO will handle, clean out template instructions except for those sections still blank.	0.64
Lombardo	4/5/05	Fleshed out OCIO input areas	0.65
Sadler	4/5/05	Added user class for non-ARS researchers, fixed typos in section 2.7 and wherever 'thee' was, and rearranged section 3.6-3.9 into sections 3.6.1-3.6.4, although it is not a little hard to group the data retrieval with the data uploading.	0.66
Wilson	4/7/05	Modified 3.1.2, 4.1, 4.2, 4.3, 4.4 and added appendix E (Sadler accepted some changes and updated Table of contents and this table)	0.67
James	4/11/05	Add to References (section 1.5). Add to Analysis Models (Appendix B)	0.68
Sadler	4/13/05	Performed overview edit for internal consistency, grammar, typos, and some aspects of 12 vs more watersheds.	0.69
Lombardo	4/19/05	Security Requirements Section rewritten	0.70
Philpot	4/21/05	Formatting changes	0.80
Lombardo	4/25/05	Added clarification regarding scope of 1.0 version to Document Conventions section.	0.90
Philpot	5/4/05	Formatting and JLS edit inclusion	0.91
Greg Willson/Jin-Song Chen	2/7/06	Added System Design (database design and User Interface Design), Appendix F, G, H, I	1.00
Chen and Steiner	2/15/06	Added text in system design Appendix F, G, H, I	1.1

1. Introduction

1.1 Purpose

The USDA and Agricultural Research Service (ARS) have supported hydrologic research since the 1930's and many of the research watersheds are still operational. However, data have been managed and made available independently at each research location, greatly reducing the accessibility and utility of these data for policy-relevant, multi-site analyses. The ARS is developing and implementing a data system to organize, document, manipulate and compile water, soil, management, and economic data for assessment of conservation practices. While primary responsibility for data will continue to reside at individual watersheds, the new data system will provide one-point access to data from the sites in well-documented, and standardized formats. The purpose of this document is to describe the technical and operation requirements that meet the needs of the CEAP Watershed Assessment Studies, as well as additional needs of researchers at the watershed sites and diverse outside users, in adequate detail to provide the basis for the system design. The system will be called STEWARDS: Sustaining the Earth's Watersheds – Agricultural Research Data System.

1.2 About this Document and its Readers

The system requirements specification document describes what the system is to do, and how the system will perform each function. The audiences for this document include the system developers and the users. The system developer uses this document as the authority on designing and building system capabilities. The users review the document to ensure the documentation completely and accurately describes the intended functionality.

This version – version 1.0 - provides general descriptions of the system. The system developer should review the document to ensure there is adequate information for defining an initial design of the system. The users should review the document to affirm the features described are needed, to clarify features, and to identify additional features needed within the system.

The next version – version 2.0 – will be the result of more detailed requirements analysis. When version 2.0 is written, the system developer and users will be asked to review this document.

The document is structured to follow IEEE 830-1998 standards for recording system requirements.

1.3 Project Scope

As part of the Conservation Effects Assessment Project (CEAP), twelve ARS Benchmark Watersheds will support watershed-scale assessment of environmental effects of USDA conservation program implementation. The ARS Benchmark Watersheds represent primarily rainfed cropland, although some of the watersheds also contain irrigated cropland, grazingland, wetlands, and confined animal feeding operations. Three additional ARS watersheds may be added in the future to represent additional land uses. Conservation practices to be emphasized will include NRCS CORE 4 practices for croplands (conservation buffers, nutrient management, pest management, and tillage management), drainage management systems, and manure management practices. Environmental effects and benefits will be estimated primarily for water and soil resources, with some assessment of wildlife habitat and air quality benefits in some watersheds.

The goal for the watershed-scale research is to provide detailed assessments of conservation practices and programs in a selected watershed, provide a framework for improving the performance of the national assessment models, and support coordinated research on the effects of conservation practices across a range of resource characteristics (such as climate, terrain, land use, and soils).

The basic requirement for STEWARDS is to deliver consistent high quality data in support of the CEAP watershed-scale assessment from a one-stop data portal to the CEAP clients and, in time, the public at large. The data system consists two main parts -- a central database management system for the storage and management of data, and a client application to allow users access and interact with the data. The data stored in the database can be viewed and downloaded using graphical and tabular interface tools.

The data system will serve as a repository where diverse end-users can access, search, analyze, visualize, and report various types of integrated watershed data contributed from the multiple locations. Types of data will be diverse, including biophysical data (i.e., point-based in time and/or space, spatially variable data, time series), data about land use, management, and conservation practices; and economic data. Where applicable, data from the NRCS, Economics Research Service (ERS) Agricultural Resources and Environmental Indicators database, other ERS sources of data on costs of conservation practices, and land use and management data of NASS or other agencies will be utilized.

The intention of the team is that ARS researchers at watershed locations retain primary responsibility for data collection and management. Sites may chose to retain existing data management protocols or change location protocols to those of STEWARDS. However, STEWARDS must have the capability to translate data from location-specific formats into the standardized STEWARDS formats. Data on STEWARDS will not be available to the public in real time. Access to real-time data will remain at the discretion of each location research team. Instead, annual or more frequent updates from location databases to STEWARDS will be made by the watersheds' staffs after the locations have completed their quality assurance procedures.

1.4 References

Conservation Effects Assessment Project:
http://www.nrcs.usda.gov/technical/nri/ceap/

Content Standard for Digital Geospatial Metadata (version 2.0), FGDC-STD-001-1998:
http://www.fgdc.gov/metadata/contstan.html

USDA Metadata Tool for creating SCI minimum compliance metadata:
http://www.gis.sc.egov.usda.gov/software/tools/arcgis/metadata-tool.html

Soil and Water Assessment (SWAT) Model:
http://www.brc.tamus.edu/swat/edu.html

AnnAGNPS Pollutant Loading Model:
http://www.ars.usda.gov/Research/docs.htm?docid=5222

USDA Web Style Guide
http://sharedservices.usda.gov/

IEEE 830-1998

Institute of Electrical & Electronics Engineers Recommended Practice for Software Requirements Specifications

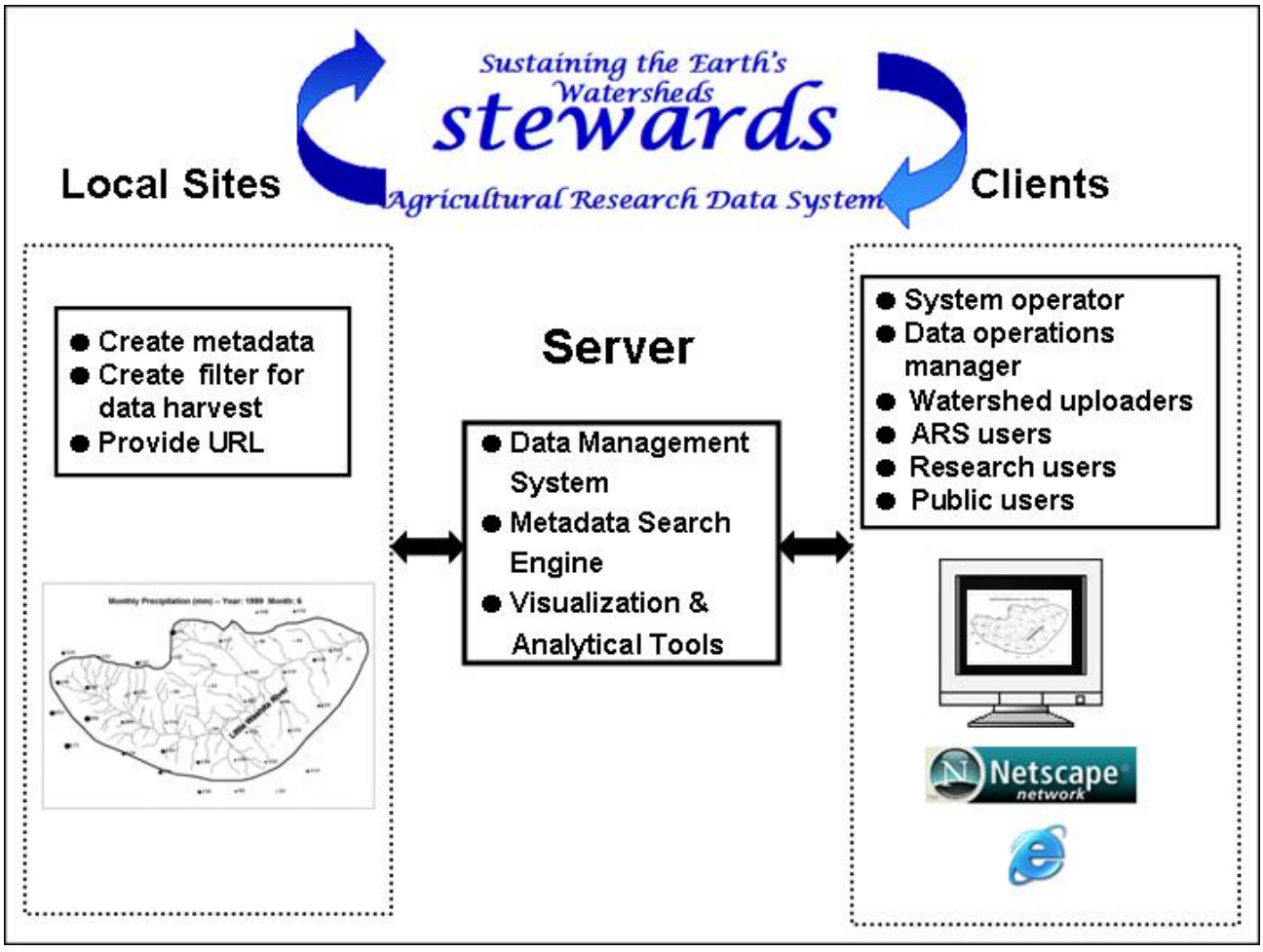
2. Overall Description

2.1 Product Perspective

This product is a new, centralized data system. Source data for this system are the long-term ARS watersheds, with those participating in the CEAP project expected to participate in this data system first.

2.2 Product Features

The data system consists two main parts -- a central database management system for the uploading, storage, and management of data, and a client application to allow users access and interact with the data. Diverse end-users can access, search, analyze, visualize, download, and report various types of integrated watershed data contributed from the multiple locations. Types of data will include biophysical data (i.e., point-based in time and/or space, spatially variable data, time series), data about land use, management, and conservation practices; and economic data. See figure 1.



2.3 User Classes and Characteristics

The user classes of this database system will be

UCC-1: System operator: Staff at OCIO responsible for maintaining executable code, requiring full access to all aspects of the system.

UCC-2: Data operations manager: Staff, probably at OCIO, responsible for data archiving and maintenance, requiring read/write/modify privileges to all data in the system. This includes the role commonly thought of as a Database Administrator (DBA).

UCC-3: Watershed uploaders: Staff in the watersheds who have responsibility to upload local data requiring write access to space allocated to their specific watershed.

UCC-4: ARS users: ARS scientists, engineers, and support staff who require ARS watershed data for research purposes. These would likely be conversant with the general characteristics of the data content, and to a lesser extent have abilities manipulating views of the data using database tools such as envisioned for this database system. These users will need user/password authentication to enable user settings management and to have access to data protected through the confidentiality agreements among agencies participating in CEAP.

UCC-5: Research users: Non-ARS scientists, engineers, and support staff who require ARS watershed data for research purposes. These would have similar characteristics and needs as ARS users, but because of the agency-level confidentiality issues, could not access the password-protected sensitive data.

UCC-6: Public users: Eventual release of the database system to the general public will add university and agency researchers who would appear similar in training to the primary users, but also would add clients with the entire range of training and interests. These users would have access to the public data without need for authentication.

2.4 Operating Environment

The rule for selecting hardware and software is that the components/application must be functionally efficient, capable of interfacing with other software, and easy to maintain.

OE-1: The System shall operate with the following Web browsers: Microsoft Internet Explorer versions 5.0 and 6.0, Netscape Communicator version 4.7, Netscape versions 6 and 7, and FireFox.

OE-2: The System shall operate on Certified and Accredited servers running the current corporate approved versions of operating systems as appropriate.

OE-3: The Data Management portion of the system shall permit user access from the corporate Intranet and, if a user is authorized for outside access through the corporate firewall, from an Internet connection at the user's home.

OE-4: The Data Research and Reporting portion of the system shall provide public access to reviewed and approved data sets.

2.5 Design and Implementation Constraints

The goal of the web application is to be platform independent on the client side wherever possible. Therefore, the web applications will be implemented to run on the server side as much as possible. Also, it is required to test the application using different platforms, connection speeds, screen settings, colors/graphics, and browsers.

CO-1: Portability: The application codes generated during prototyping may not run properly when re-hosting the data system to a non-Windows OS or transferring the data system to another location.

CO-2: Database Design: The database structure should be as complete as possible during the design stage but there should be a room for modification without a large overhaul during later phases.

CO-3: Data uploading and management will likely be done as infrequently as once a year. This fact greatly limits the complexity of the user interface and the learning curve needed for completing this task.

CO-4: The system shall use the current corporate standard Microsoft SQL Server database engine.

CO-5: The system shall be in compliance with all Accessibility, Web Design, and Security policies applicable.

CO-6: As part of standard operating procedures, a testing plan will be documented during the Design phase. The testing plan will be based on user roles, modules or use cases, required tasks and expected outcomes.

2.6 User Documentation

User documentation will consist of the several components usually expected of a modern web-based software application, including a tutorial, help pages, FAQ's with an online request form, and a complete user's manual.

UD-1: A tutorial will provide a quick start, a walk-thru of major system features, and further reference source for the complete system features.

UD-2: An online hierarchical and cross-linked help system in HTML will describe and illustrate all system functions.

UD-3: An online form will enable users to request help. Frequently asked questions will be screened for the FAQ pages.

UD-4: The user's guide (or user manual) will contain sufficient information and instructions required to access and use the data system. It will include:

UD-4.1. Overview of the system features and architecture.

UD-4.2. Instructions for accessing the system.

UD-4.3. Samples of screens, where appropriate.

2.7 Assumptions and Dependencies

It is anticipated that each watershed location will provide human and fiscal resources required to prepare metadata files, data files, work with the data team on initial import of data, perform subsequent data uploads, and provide site-specific support to users of the system. If a location has inadequate resources, the minimal data required for CEAP analyses will be identified and the data team will provide additional support to that location. This may result in a location's data coming into the system later than scheduled.

Team leaders anticipate availability of the operational platform at the ARS OCIO level. Pilot and, as a contingency, operational data systems may be developed and maintained at El Reno or other research locations.

It is anticipated that funding from NRCS will partially support this activity through FY07. If that funding were not available or insufficient in future years, base funds at the locations and/or discretionary funds from the Area and Headquarters funds would support the activity but the timelines would be adjusted.

Help desk funding will be determined from staffing needs identified during the design phase and a proposal will then be made to NPS.

3. System Features

3.1 Database Management System

The standardized data structure will facilitate database development, storage, and maintenance; metadata development, storage, and maintenance; and support database functionality, including tabular and spatial data queries, evaluation, visualization, and transfer (downloading) to off-site locations.

3.1.1 Description and Priority

The database management system should be stable and secured, have high performance (in terms of speed and efficiency/effectiveness), and be easy to maintain. This component is central to the effort and is thus of highest priority.

3.1.2 Use Case Scenarios

The team will elicit user requirements from a cross-section of user class members (see section 2.3) to accurately and completely describe the expected uses and functionality of the system. To elicit requirements, the user class members will be asked via conferencing, interviews, and email, to provide a detailed description of the user actions and system responses that will take place during execution of the use case under normal, expected conditions. The summation of responses will lead to an accurate and complete inventory of user requirements.

3.1.3 Functional Requirements

FR-1: DATABASE

FR-1.1: Database design

Establish a database in support of CEAP that will house the collective data assembled or generated during research activities. The database will support a variety of data types and formats, including but not limited to: spatial data - vector, raster, imagery, and tabular; tabular data – static and time-series; spreadsheets; documents; reports; photographs; and URL links.

FR-1.2: Modeling support

The use of agricultural models (SWAT, AnnAGNPS) is a core element of the CEAP effort. Several modeling-related database topics will need to be addressed:

1. Maintenance and reporting of uncertainty or error in measured data, in spatial components of GIS data, and in modeling output. Reporting of uncertainty may take place on an individual sample basis, on a method or procedural basis, or on an entire database.

2. Models will require specific input data and will generate output data during the calibration and validation phases, sensitivity analyses, and exploratory scenarios. The current scope of the STEWARDS system is to provide measured data for input to the model and for comparison to output.

FR-2: DATA ACCESS

FR-2.1: View entire universe of watersheds from top-level screen for selection

Design tools to navigate the individual watershed sites and their data.

FR-2.2: View watershed descriptive data

Summary descriptions of research watersheds, stations, and instruments that are useful in describing the research activities should be accessible. This information would not be considered 'formal' metadata.

FR-2.3: Browse, query, and download individual sampling station data and metadata

Provide access to the data via browsing of sites, stations, and instruments; allow for simple queries to individual datasets; provide a metadata search tool to query dataset parameters; and allow for downloading of datasets (full or partial).

FR-2.4: Visualize time-series data

Users may wish to examine time-series data, e.g. stream discharge data, over a user-specified time frame in order to select only those data desired for download. Charting tools in association with the query engine are desirable.

FR-2.5: Generate tabular reports of selected data

Provide access to CEAP-related reports, tables, and project documents.

FR-2.6: Visualize, query, and download spatial data

Agricultural research data are inherently spatial in nature. STEWARDS will provide web-based geographic information tools to allow site-specific data to be viewed within their spatial context. These tools will provide browse and query functionality and support links to download spatial data and their associated tabular data.

FR-2.7: Browse, visualize, and download agricultural modeling data and results

Much of the initial effort in CEAP will focus on the use of agricultural models (e.g. SWAT, AnnAGNPS) in the CEAP watersheds. Users will be able to examine the data used in the modeling effort, visualize the results in their spatial context, and download the model input data and results.

FR-3: METADATA

FR-3.1: Maintain metadata database

Develop and maintain a metadata database that provides browse and query access to formal descriptions of the database elements. Provide tools for database contributors to create and upload metadata compatible with the STEWARDS metadata database.

FR-3.2: Browse and query watershed metadata by theme or location

Develop search tools that query watershed, station, and instrument metadata across the STEWARDS database. Metadata query tools will support query by location, theme, and keyword.

FR-4: SYSTEM SUPPORT

FR-4.1: Administrative metrics – user access lists, data download estimates

FR-4.2: User feedback, user support tools

FR-4.3: Conform to USDA web site style guide

3.2 Metadata Availability

Metadata are data about data. The STEWARDS database will maintain a metadata report for each database in STEWARDS. The metadata of individual datasets should be created at the local watershed sites for local use and for populating a CEAP metadata database. Elements of the STEWARDS Metadata system will include:

MA-1: Metadata standard
Adopt a metadata standard that is compliant with Federal regulations. The current Federal standards for spatial data are the *Content Standard for Digital Geospatial Metadata (version 2.0),* (http://www.fgdc.gov/metadata/contstan.html), FGDC-STD-001-1998. There are some indications that the Federal government may implement the ISO 19115 geographic information/metadata standard in the future. If this occurs, the CEAP database team will evaluate conversion of metadata.

MA-2: Metadata input tool
Implement a metadata input tool that supports creating and editing of metadata in the chosen format. The metadata tool will allow for local and web-based input and editing. A likely scenario involves a web-based wizard which allows users to specify particulars of a data set, and then save those answers and relate them to a particular project, so that these selections can be re-used in future uploading operations.

MA-3: Metadata query system
Implement a web-based browse-and-query tool for searching STEWARDS metadata. Queries by location, theme, and keyword will be supported. A likely scenario involves a web-based wizard which allows users to specify particular querying parameters and then save them for reference later such that time-based comparisons are facilitated with minimal user training.

MA-4: Metadata database
Implement a database of STEWARDS metadata elements. The database will provide browse and query functionality.

3.3 Data Dissemination

To make data electronically available any place and any time through a web server-client service, the web service will be platform independent as much as possible. Where there is an exception, the software/hardware requirement should be specified for a particular application. Users whose computing environment precludes downloading the data via the internet will have the opportunity to acquire data, reports, model results and other STEWARDS information via alternative media.

3.4 Hardware requirements

HR-1: Data Storage

HR-1.1: Online Storage - The system needs to be able to store hundreds of megabytes of data on demand. The potential for needing Gigabytes of data storage capacity is also in the realm of possibilities. Further requirements gathering is needed to get real-world estimates of such data storage needs.

HR-1.2: Near-line Storage – All User and Application data as well as software installation and configuration files must be fully backed up week-nightly. Further, secure, off-site storage will be performed on a weekly basis with 24-hour retrieval times.

HR-2: Networking

HR-2.1: Load Balancing - Application Servers and Database servers must be load balanced at the application level to ensure maximum stability and availability. Specific scenarios, such as fail-over versus session-managed load balancing need to be addressed in the functional requirements specification.

HR-2.3: Private Network - The system requires a “back-end” private network environment in order to ensure that end-user operations (on the front-end) are not impeded by database backups, index propagations, or other large back-end data transfers.

HR-2.4: The system will use, where appropriate, the standard hardware and data communications resources provided by the ARS OCIO at the ARS George Washington Carver Center in Beltsville, MD. This includes, but is not limited to, the general Ethernet network/T1 connection at the server/hosting site, network servers, and network management tools.

3.5 Software requirements

SR-1: Backup Software – Data and application backups will be managed through fully supported backup software solutions.

SR-2: HTTP and GIS Server Applications – As the Web will be the primary delivery protocol for the application, HTTP and related GIS server applications will be required to support system functionality.

SR-3: Web Browsers and Browser Plug-ins – In support of External Interface requirements, commonly supported web browsers will be used to implement a thin-client architecture. The use of Browser plug-ins will be judiciously restricted on an as-needed basis.

SR-4: Email Services– As a secondary delivery protocol for alerts, data, and other information from the system, email server applications will be required to support system functionality.

SR-5: Relational Database Management System - As the primary data storage mechanism for the corporate standard relational database management system, Microsoft SQL Server will be required to support system functionality.

SR-6: The system will use, where appropriate, the standard software resources provided by the ARS OCIO. This includes, but is not limited to, MS ASP/Java Scripts, $C^{++}/C^{\#}$, MS SQL server and IIS, or the use of PHP/Perl, JB scripts, $C^{++}/C^{\#,}$ MySQL server, and Apache server.

SR-7: The system will make data available in standardized data formats such as tab-delimited text for use on users' client systems including local installations of ARC GIS software or other off-line software applications.

3.6 Data migration to and retrieval from STEWARDS

Data will be subject to quality assurance prior to either initial or recurring uploading to STEWARDS. Initial population of STEWARDS will require that transition filters be developed jointly between the STEWARDS team and the individual watersheds. These filters will provide the same function for recurring, probably annual, uploads.

3.6.1 Quality Assurance

Local watershed staff will be responsible for performing necessary quality assurance/quality control (QA/QC) procedures prior to data being uploaded to the STEWARDS database. Measurement methods and QA/QC procedures established by the CEAP Quality Assurance Team will be used by local data producers and recorded in the data provided to STEWARDS.

The purpose of quality assurance is intended to provide added assurance of data integrity. Quality assurance checking will be carried out at two different levels – local site level and central site level. First, the guidelines for quality assurance, including parameter-specific ranges, threshold values, quality flags etc., will be determined. Quality assurance checking at the local sites may be performed before or during the standard exchange file conversion phase. A general quality assurance checking, not as specific as at the local sites, will be carried out during the data uploading stage. Note that a) the primary responsibility for data quality assurance rests with the individual sites and b) the quality assurance ranges, criteria, etc., used here should be consistent with the quality assurance information specified in metadata (and may be a component of the data dictionary).

3.6.2 Populating data

Basic information regarding individual watersheds will be made available to STEWARDS by local watershed sites. Basic information may include spatial data in a GIS-compatible format; descriptive data about stations, instruments, methods, or procedures; reports; modeling results; and other data. Data will be reformatted to be compatible with the STEWARDS data schema and CEAP modeling requirements. Watershed data to be used in modeling that does not conform to model requirements will be processed in STEWARDS for inclusion in the STEWARDS databases.

As part of the initial population of the database, time-series data (measurement data, annual land use information, weather/climate data) that has been collected by local watershed staff will be processed through local and CEAP QA/QC procedures and uploaded to the STEWARDS database by watershed staff per obligations to CEAP Objective 1.

Protocols for data imports from local sites via input files with different formats will be developed and implemented jointly between the STEWARDS team and the local watersheds. The standard exchange file format will have been jointly determined first. A single import filter will be written to parse the standard formatted (e.g. tab-delimited) files from the local watershed sites. In this manner, all the data files from the local sites will have consistent formats when they are stored in the central site. The filter developed for the intial population of the database will be available for use in later periodic uploads.

3.6.3 Updating Data

Time-series data (measurement data, annual land use information, weather/climate data) and other data (spatial data, metadata) will be collected by local watershed staff, processed through local and CEAP QA/QC procedures, and uploaded to the STEWARDS database on an annual basis using the filter developed during the initial population of the database. Requests will initiate from the central system per schedule negotiated with individual watersheds. Compliance by watershed staff is understood to be expected per obligations under CEAP watershed project plans revised to accommodate the CEAP project plan objective 1.

3.6.4 Data derivation/aggregation

The central database will allow users to access the data of individual watersheds, sub-watersheds on a station-by-station basis, or a collection of stations if the entire dataset is desired. Multiple levels of data aggregation of time-series data (daily, monthly, or yearly) will be supported. A great flexibility of export data should be also available in terms of data volume (the number of data entity and time coverage) and output format [for example, for tabular data – comma separated values (csv) or database format (dbf); for spatial data – shape files (shp) or ascii grid files (grd); for reports – Adobe page definition format (pdf)].

4. External Interface Requirements

4.1 User Interfaces

The user interface will be simple and consistent, using terminology commonly understood by the intended users of the system. The system will have a simple interface, consistent with industry standard interfaces, to eliminate the need for user training of infrequent users. The STEWARDS team will evaluate the user interface of similar systems and apply appropriately. For additional details see Appendix E. User testing will be used to ensure the user interface is clear (simple, commonly understood vocabulary, intuitive to use without training), complete (users can perform all functions from the interface), and consistent (buttons and wording are the same throughout the system).

4.2 Hardware Interfaces

No extra hardware interfaces are needed. The system will use the standard hardware and data communications resources provided by the ARS OCIO at the ARS George Washington Carver Center in Beltsville, MD. This includes, but is not limited to, the general Ethernet network/T1 connection at the server/hosting site, network servers, and network management tools. This system will include a warning message when a low transmission speed is detected, and a non-graphical interface option will be available.

4.3 Software Interfaces

The system will use the standard software resources provided by the ARS OCIO. This includes, but is not limited to, MS ASP/Java Scripts, C^{++}/$C^{\#}$, MS SQL server and IIS, or the use of PHP/Perl, JB scripts, C^{++}/$C^{\#,}$ MySQL server, and Apache server.

4.4 Communications Interfaces

The system will use the communications resources provided by the ARS OCIO. This includes, but is not limited to, HTTP protocol for communication with the web browser and the web server and TCP/IP network protocol with HTTP protocol.

5. Other Nonfunctional Requirements

5.1 Performance Requirements

PR-1: Response time: The data system shall show no visible deterioration in response time as the number of persons increases. Response times seen by end users for querying metadata should be on the order of a few seconds or less. Response times seen by end users for retrieving the actual images may take much longer, anywhere from a few minutes to several hours. If the user requests a large image with a short response time, it is acceptable for the ARS data system to advise the user that the target response time cannot be met. In that case, the person would be referred to an alternate method of getting the data.

PR-2: Loading speed: The data system shall load as quickly as comparable productivity tools on whatever environment it is running in.

5.2 Safety Requirements

Data on the server should be protected from power loss but data in transit from server to requester could be lost. Given that these data will also remain on the watershed site system, rather than expend resources to prevent this loss, such failures will be monitored and the uploading process will be repeated.

5.3 Security Requirements

STEWARDS will have reasonable controls consistent with ARS OCIO practices, in compliance with agency, Dept and Govt regulations. STEWARDS security requirements will have four primary components. They are authentication, confidentiality, integrity, and availability.

SCR-1: Authentication
STEWARDS will follow industry best practices for authentication, using single-sign-on systems like Microsoft Active Directory to perform authentication. Authentication addresses security requirements to ensure those using system are who they say they are. This is of greatest concern when data are being changed or updated. This is primarily done through userids and passwords.

SCR-2: Confidentiality

Confidentiality security requirements describe the need to protect the data appropriately. STEWARDS will use the user classes described in section 2.3 above to define boundaries of information sharing to ensure confidentiality as appropriate. Any data that should be viewed by a restricted audience must be protected with appropriate security features.

SCR-3: Data Integrity

The integrity of STEWARDS data will be critical to its success as a product. Scientific research and publications will be based on the data obtained through the system. Therefore, extensive data validation and review will be performed both before data are uploaded to the system and as part of the upload process. The system will need policy and procedures protecting the data from intentional or unintentional modifications, and to ensure accurate data are made available.

SCR-4: Availability

The fourth consideration for security requirements is availability. The system must be available to the intended audience 24 hours per day, 7 days a week with, 99% availability and a tolerance of -5% (not less than 50% of working hours in any week). For this system, availability will be concerned with the reliability of the software and network components. Intentional "denial of service attacks" is not foreseen as a significant concern.

5.4 Software Quality Attributes

SQ-1: Portability

This database will be built for a particular system and may not be portable but results to queries will be portable between many environments.

SQ-2: Adaptability

Implementation of the application software/code and design of database structure should be flexible enough for the necessary change in the later phase.

SQ-3: Availability

Availability is defined here to mean the ability to use the system during its intended period of operation as defined in SCR-4 above.

Appendix A: Acronyms and Glossary

Acronyms:

AnnAGNPS - Annualized Agricultural Non-Point Source
ARS - Agricultural Research Service
CEAP - Conservation Effects Assessment Project
DBA - Database Administrator
ERS - Economics Research Service
GIS - Geographical Information System
HTML - Hyper Text Markup Language
HTTP - HyperText Transfer Protocol
IIS - Internet Information Service
ISO Standards - Standards established by International Organization for Standardization
MS ASP - Microsoft Active Server Page
MS SQL server- Microsoft Structured Query Language Server
NRCS - Natural Resources Conservation Service
OCIO - Office of Chief Information Officer
Perl - Practical Extraction and Report Language

QA/QC - Quality Assurance/Quality Control
STEWARDS - Sustaining the Earth's Watersheds – Agricultural Research Data System
SQL - Structured Query Language
SWAT - Soil and Water Assessment
URL - Uniform Resource Identifier
USDA - US Department of Agriculture

Glossary:

AnnAgNPS -- see Appendix B

Apache - An open-source web server provided by the Apache Software Foundation

ArcGIS - ArcGIS is an integrated collection of GIS software products for building a complete GIS developed by ESRI.

Authentication - The procedure (essentially approval) used by the approval authority in verifying that specification content is acceptable. Authentication does not imply acceptance or responsibility for the specified item to perform successfully.

Benchmark watersheds - is used to differentiate the larger scale ARS watersheds from field scale ARS research activity or from other non-ARS watersheds.

C++ - "C plus plus" is a programming language that was built off the C language.

C# - A modern object-oriented programming language designed and developed by Microsoft

Client – (1) A computer process that requests a service from another computer and accepts the server's responses; (2) the individual computers in a network computing system

Core 4 - NRCS's Core 4 is a common-sense approach to improving farm profitability while addressing environmental concerns. The approach is easily adaptable to virtually any farming situation and can be fine tuned to meet your unique needs. The net result is better soil, cleaner water, greater on-farm profits, and a brighter future for all of us.

Data aggregation - Any process in which information is gathered and expressed in a summary form, for purposes such as statistical analysis

Database - A collection of related data stored in one or more computerized files in a manner that can be accessed by users or computer programs via a database management system.

Database management system - An integrated set of computer programs that provide the capabilities needed to establish, modify, make available, and maintain the integrity of a database.

Data dictionary - A collection of data definitions, each identified by the name of a piece of data, describing the meaning and purpose of that piece of data in the system, the type of the data, its components and any other relevant attributes (range, precision, storage size and so on).

Data Migration - The process of translating data from one format to another or from one storage device to another

Data Schema - A data schema is a grammar that describes elements and their types, attributes and possible relations.

Functional requirement: A statement of a piece of required functionality or a behavior that a system will exhibit under specific conditions. These include inputs, outputs, calculations, external interfaces, communications, and special management information needs. Functional requirements are also called behavioral requirements because they address what the system does.

Intranet - An intranet is an internal or private Internet used strictly within the confines of a company, university, or organization.

JavaScript - A programming language designed by Sun Microsystems, in conjunction with Netscape, that can be integrated into standard HTML pages.

Metadata -- "data about data" describe the content, quality, condition, and other characteristics of data.

Module – (1) In software, a module is a part of a program. A single module can contain one or several routines. (2) In hardware, a module is a self-contained component.

MySQL - A public domain, open-source database system using SQL

Performance - A quantitative measure characterizing a physical or functional attribute relating to the execution of a mission/operation or function.

Performance requirement -- A system/software system requirement specifying a performance characteristic that a system/software system or system/software component must possess; for example, speed, accuracy, and frequency.

PHP - An HTML-embedded Web scripting language

Populating Data – initial uploading of data into the database

Portability - (1) A term used to describe an object that can be easily moved, such as a portable computer; (2) When referring to computer software, portability refers to how easy a software program can be moved between computer Operating Systems.

Requirement -A statement of need for some aspect of a system, often elicited directly from a stakeholder or captured from a source document

Server – A central computer (server) which provides services such as file storage, printing, and communications in a network computing system

Software requirement – (1) A software capability needed by a user to solve a problem to achieve an objective; (2) A software capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document.

SWAT - see Appendix B

System - A composite of equipment, skills, and techniques capable of performing or supporting an operational role or both. A complete system includes all equipment, related facilities, material, software, services and personnel required for its operation and support to the degree that it can be considered a self-sufficient item in its intended operational environment.

System Requirement - A condition or capability that must be met or possessed by a system or system component to satisfy a condition or capability needed by a user to solve a problem.

Use cases - A task analysis technique often used in software engineering. For each module of a system, common tasks are written up with the prerequisites for each task, the steps to take for the user and the system, and the changes that will be true after the task is completed. Use cases are especially useful for making sure that common tasks are supported by the system, that they are relatively straightforward, and that the system architecture reflects the task structure.

User class - A group of users for a system who have similar characteristics and requirements for the system

User interface – A user interface is what you have to learn to operate a machine. For examples, the graphical user interfaces (GUIs) -- windows, icons, and pop-up menus have become standard on personal computers.

User requirements - address what the users need to do their jobs. These requirements are implementation independent and are sometimes called "business requirements." Read about the important of user requirements.

Appendix B: Analysis Models

SWAT - Soil and Water Assessment Tool, a river basin, or watershed, scale model developed by Dr. Jeff Arnold for the USDA Agricultural Research Service (ARS). SWAT was developed to predict the impact of land management practices on water, sediment and agricultural chemical yields in large complex watersheds with varying soils, land use, and management conditions over long periods of time.

AnnAgNPS - AGNPS is a tool to help evaluate the effect of management decisions impacting a watershed system. The AGNPS system is a direct update of the AGNPS 98 & 2001 system of modules containing many enhancements. The term "AGNPS" now refers to the system of modeling components instead of the single-event AGNPS, which was discontinued in the mid-1990's. These enhancements have been included to improve the capability of the program and to automate many of the input data preparation steps needed for use with large watershed systems. New to AnnAGNPS Version 3.42 are many minor enhancements to algorithms and more output options. The AGNPS Arcview interface has been better integrated with the components needed to develop AnnAGNPS datasets. The capabilities of RUSLE, used by USDA-NRCS to evaluate the degree of erosion on agricultural fields and to guide development of conservation plans to control erosion, have been incorporated into AnnAGNPS. This provides a watershed scale aspect to conservation planning.

Appendix C: Issues List

This section is reserved for open requirements issues that remain to be resolved, including TBDs, pending decisions, information that is needed, conflicts awaiting resolution, and the like.

Appendix D: Database team members

Steiner, Jean	El Reno, OK	405-262-5291	jean.steiner@ars.usda.gov
Sadler, John	Columbia, MO	573-884-1971	sadlerj@missouri.edu
Anderson, Dave	Ft. Collins, CO	970-295-5539	danderson@itc.nrcs.usda.gov
Batten, H. L.	Tifton, GA	229-386-3297	hlb@tifton.usda.gov
Chen, Jin-Song	El Reno, OK	405-262-5291	jin-song.chen@ars.usda.gov

Herring, Rob	Beltsville, MD	301-504-1050	rherring@ars.usda.gov
James, David	Ames, IA	515-294-6858	james@nstl.gov
Lombardo, Pete	Beltsville, MD	301-504-1073	pvl@ars.usda.gov
Philpot, Jill	Beltsville, MD	301-504-5683	jphilpot@ars.usda.gov
Ross, John	El Reno, OK	405-262-5291	john.ross@ars.usda.gov
Wilson, Greg	Beltsville, MD	301-504-1058	gwilson@ars.usda.gov

Appendix E: System Detailed Specifications

DS-1: General
- GUI and dynamic web applications will be used as much as possible in the ARS data System.
- Project-specific standards will be used to ensure consistency across the tools. The pages can be viewed with Microsoft Internet Explorer versions 5.0 and 6.0, Netscape Communicator version 4.7, Netscape versions 6 and 7, and FireFox.1.0.
- The pages will be designed for screen resolution of 800 x 600 pixels as a minimum.
- Proper functionality of the application will require that cookies be enabled in the browser settings.
- Arial font will be used across the tools.

DS-2: Readability
- Terminology used in the tools should be easily understandable by the user.

DS-3: Page design
- All pages performing similar functionality should have consistent Look & Feel.
- Appropriate titles should be given to each page. The titles should specify the functionality of the page.
- Appropriate alternate text should be provided for all the images, to help in navigation and better readability.
- The title tag and the Meta description tag in the web pages will be carefully constructed to include the focused keywords so these keywords will be highly visible in the search engine results pages, making the STEWARDS product more useful to the general users after the public rollout.

DS-4: Navigation:
- Navigation facilities will be provided to navigate from one page to another page with minimum number of clicks.

DS-5: Messages:
- Descriptive and user friendly messages for invalid input should be provided in all the pages. Completion/Confirmation messages should be displayed when the application processes the data successfully. Messages generated shall be clear, succinct, and free of jargon.

Appendix F: System Design

System Design Goals:

The purpose of system design is to translate the system requirements into more technical specifications. Through a user case survey and a logical analysis of user's expectations and system functionalities, the system components for STEWARDS are identified (Fig. F1). System design provides a detailed description for each identified component and produces a physical data system design (with documentation) that allow the system developers to construct each component (layer) accordingly. At the end of the system design phase, hardware, software tools and skill requirements, and other resources will be identified and a system developer team/plan will be drawn. After that, the construction phase will begins (Task flow of can be seen in Fig. H1 in Appendix H).

System Design Methodology:

The system design process can be divided into three main phases: conceptual system design, logical system design, and physical system design. In this appendix, only conceptual system design and logical system design for STEWARDS are documented. Physical system design (how the logical structure of STEWARDS will be implemented) is still in a planning stage and will be described later. Physical design, which involves producing a description of the implementation of logical system design, will be tailored to STEWARDS and all physical considerations (including tools/software such as Microsoft SQL server, Oracle SQL server etc.) will be specific to STEWARDS.

Conceptual System Design: Build a conceptual data system for STEWARDS

As depicted in the conceptual model of STEWARDS (Fig. F12), the task underway in this phase is to define/identify system components such as:

1. hardware infrastructure component,
2. data/database component (metadata standards, measurement data),
3. interface/application component (metadata search, measurement data search, data access, visualization, data downloads, and their relationships), and
4. data uploading tool component.

The database design and interface/application design parallel each other within the system development process. The total number of system components will depend on the level of detail. For example, the data component of STEWARDS can be further identified as watershed climate, image, soils, management, pollutant source, hydrology, socio-economic, water quality data, data search, data downloads etc (Fig. F2). Description for hardware infrastructure design will be provided separately.

Logical System Design: Build and validate the logical data system for STEWARDS

During the logical system design phase, a logical data system (all system components and their relationships) will be constructed and validated. In other words, the Entity-Relationship (ER) diagrams for measurement data and user graphic interfaces (UGIs) to connect all components will be developed. For example, an entity-relationship diagram for measured climate data at South Fork watershed, Iowa, is given in Fig. F3. Climate is an entity (subject) and it becomes a table (Topic(Variable)_Table which holds variables such as min_temperature, max_temperature, etc.) in the database. The actual (basic) climate data is also an entity and has its own table (a table with

yellow background in Fig. F3). Fig. F3 also shows the relationships between tables. The tables – STEWARDS_Units_lookup_Table, [WS]_Topic_Table, and [WS]_Site_ID_Table are supporting data tables to the [WS]_Climate_Daily table. The latter is the basic data table where climate data are stored. Note that the characteristics (topic) of a theme (e.g. climate) in Fig. F3 become the fields (min_temperature, max_temperature, etc.). The data descriptors for the basic data table and supporting tables are shown in Tables F1 and F2 respectively. A generalized data flow and table relationship is given in Fig. F4.

Fig. F5. is a diagram in which the data flow and relationship of system components were identified. It provides a skeleton of system interface design. Three major system application components are identified: overview/summary, system search, and data access/visualization/downloads. For next level of detail, the system component diagram for access login component, system/summary system component, metadata search component, site summary search, component, watershed specific search component, topic search component, site specific search component, pre-existing search component, characterizing watershed GIS component, and data access/visualization component are given in Figs. F6 to F15. The descriptions for these higher level system components are given in another document (Software module design documentation_v2-2.doc at The USDA-ARS ARSNET Team 1 Sharepoint site).

Data Importing/Translation:

The tasks for importing data from the local sites into the centralized site required the following considerations.

1. Assumptions:
    a. Metadata will be compliant at time of production in the local sites – no translation required, just the upload.
    b. Therefore, translating empirical data is the only need.
    c. Translation will be local site specific and done there.
    d. A thesaurus of columns and allowed parameter ids (identification IDs) will be developed either internal or external to the translation team before acting on the below.
2. There are three types of empirical data we will need to address.
    a. Time series data (Type 1 data) with columns for different data collected at the same time. Example is flat table of weather data. A generalized data structure for this type of data is given in Fig. I1 in Appendix I.
    b. Time series data (Type 2 data), usually more sparse than the above, with an id field that explains what the parameter field contains. Example is water quality data. A generalized data structure for this type of data is given in Fig. I2 in Appendix I.
    c. GIS layers.
3. There are two types of operations we need to consider
    a. Initial population
        i. Build cross-map (mapping method) of columns and/or parameter ids
        ii. Build units conversions
        iii. Choose QA procedures
        iv. Build translate/upload procedure
        v. Perform initial upload
        vi. Perform check of completeness in central store
    b. Annual population
        i. Add new columns if needed
        ii. Upload into existing (perhaps only recently) columns/fields

iii. Overwrite old data after error identification and correction locally
iv. Perform check of completeness in central store

The mapping method allows a great flexibility in terms of data heterogeneity, table size, data type, etc.

Physical System Design: Build the specific data system for STEWARDS in terms of softwares and tools.

This phase is under planning. For software requirements, the programming skills in developing Windows and Web applications using Microsoft Visual Studio .NET (BASICS, ASP.NET, C++, C#), XHTML, XML programming languages/scripts, Microsoft VISIO 2003 under MS Internet Information Services (IIS) and Microsoft SQL 2000/2005 servers will be pursued.

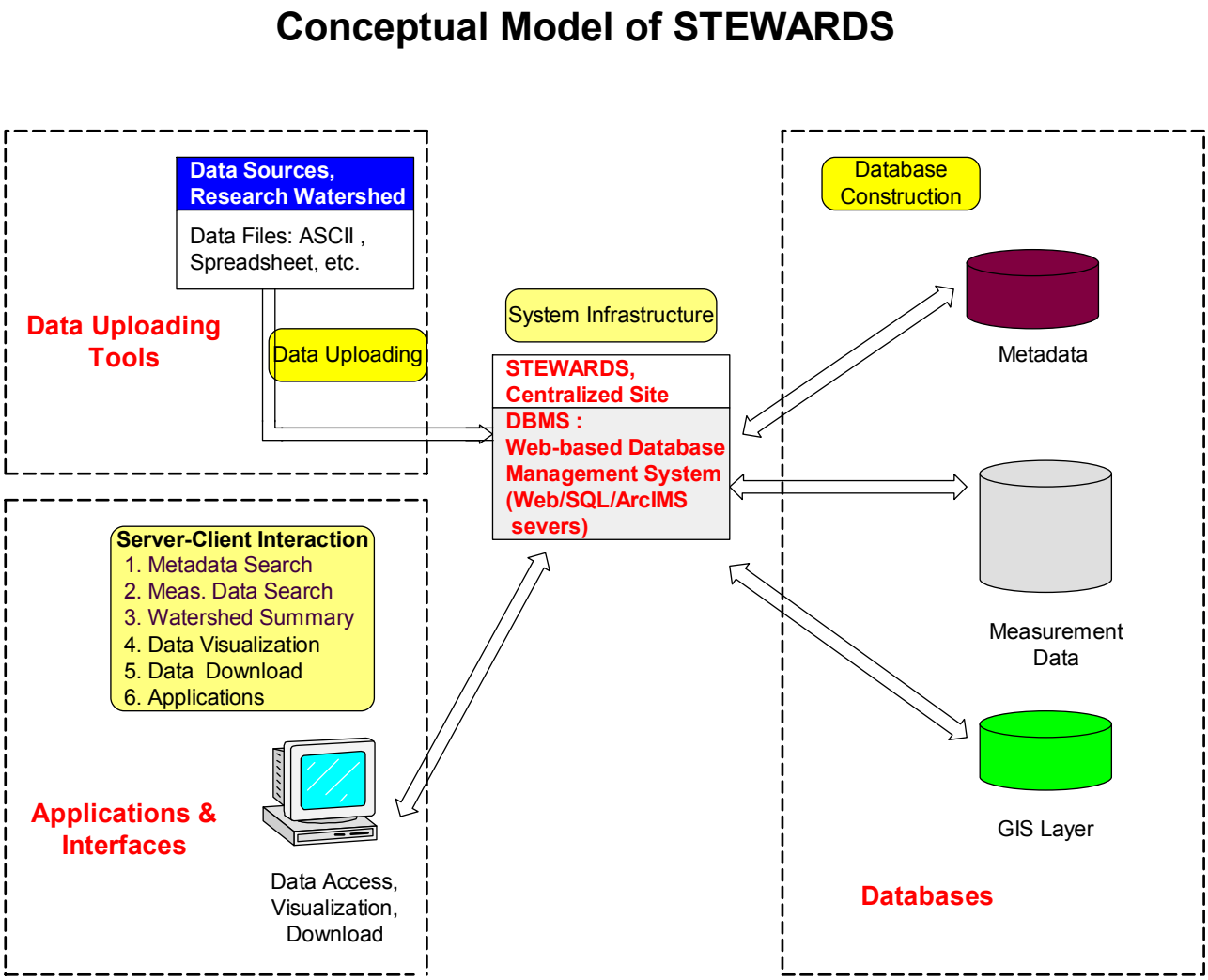


Fig. F1 Logical model of STEWARDS Data System with four system components – System infracture, databases, applications/interfaces, and data uploading.

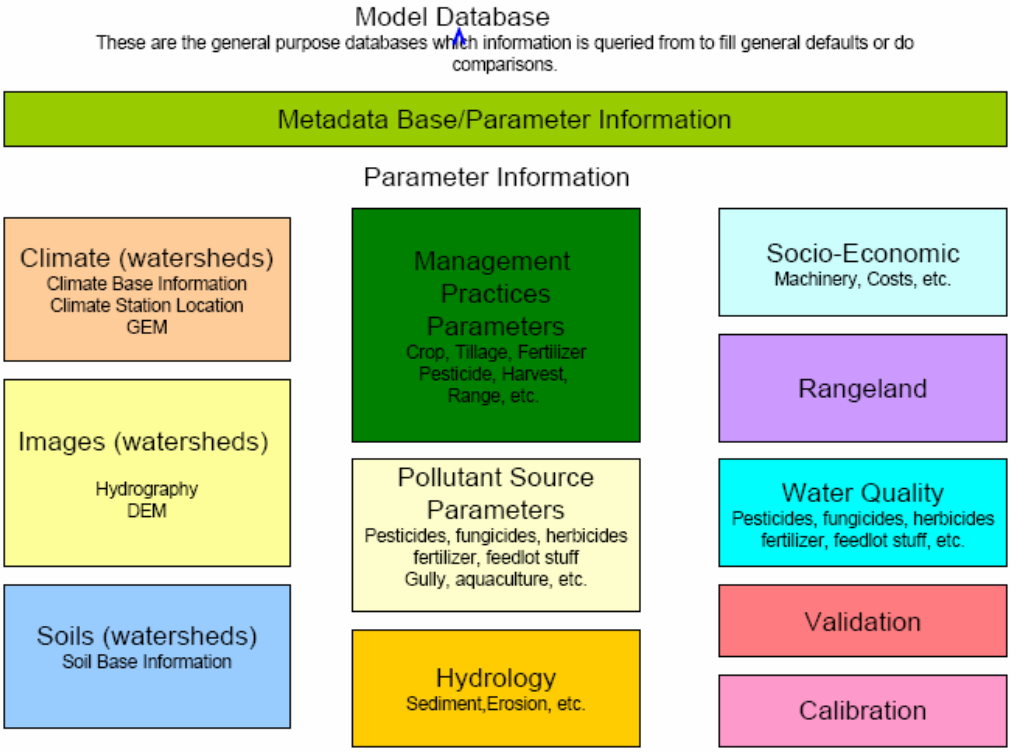


Fig. F2. Data components of STEWARDS. Each block represents a data entity (a topic object).

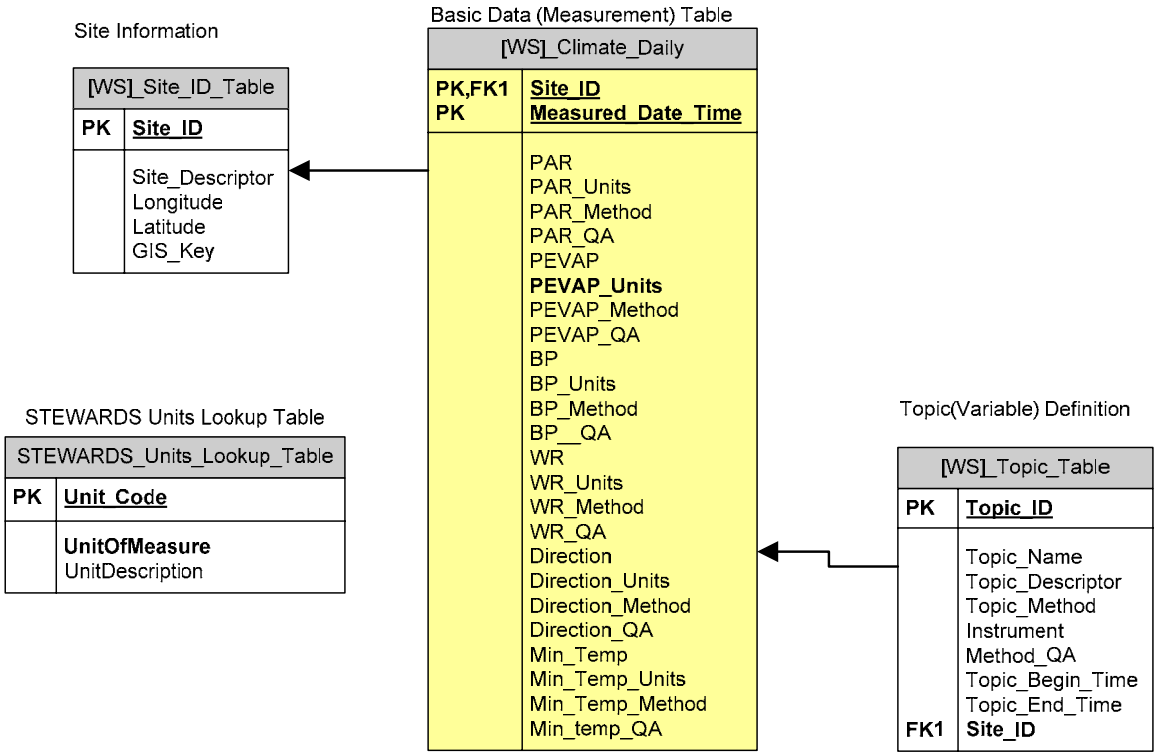


Fig. F3. An example Entity-Relationship (ER) diagram of climate data at South Fork Watershed, Iowa. The basic data table is the one in yellow color. Other tables are supporting tables The definition text at the top of the table with gray color represents the context of an entity (e.g. [WS]_Climate_Daily) while the texts under the top gray area represent the field names. PK represents primary key of a table and FK(x) is foreign key. The line with an arrow shows the relationship between two tables. A foreign key is the field in a relational table that matches the primary key column of another table. The primary key of a relational table uniquely identifies each record in the table.

Table F1. Data descriptors of measurement data table.

Measurement Data tables…

[WS]_Climate_Daily

Field Name	Description	Type and Size	Key	Relationship
Site_ID	Name of site taking measurement			[WS]_Site_ID_table
Measured_Date_Time	Year, month, day, time of day of measurement			
PAR	Photosynthetic active radiation			[WS]_Topic_table
PAR_Units	Units of measure			STEWARDS_unit_table
PAR_method	Method for measuring PAR			[WS]_Topic_table
PAR_QA	QA flag indicating measurements are within accepted standard deviation			[WS]_Topic_table
PEVAP	Pan evaporation			[WS]_Topic_table
PEVAP_units				STEWARDS_unit_table
PEVAP_method				[WS]_Topic_table
PEVAP_QA	QA flag indicating measurements are within accepted standard deviation			[WS]_Topic_table
BP	Barometric pressure			[WS]_Topic_table
BP_units				STEWARDS_unit_table
BP_method				[WS]_Topic_table
BP_QA				[WS]_Topic_table
WR	Wind run			[WS]_Topic_table
WR_units				STEWARDS_unit_table
WR-method				[WS]_Topic_table
WR_QA				[WS]_Topic_table
Direction	Average wind direction			[WS]_Topic_table
Direction_units				STEWARDS_unit_table
Direction_method				[WS]_Topic_table
Direction_QA				[WS]_Topic_table
Min_Temp	Minimum daily temperature			[WS]_Topic_table
Max_Temp	Maximum daily temperature			[WS]_Topic_table
Min_max_temp_units				STEWARDS_unit_table
Min_max_temp_method				[WS]_Topic_table
Min_max_temp_QA				[WS]_Topic_table

Table F2. Data descriptors of supporting data tables.

Supporting data tables…

[WS]_Topic_table

Field name	Description	Type and size	Key(s)	Relationship
Topic_variable	What is being measured		x	
Topic_name				
Topic_Descriptor				
Topic_Method	How data were measured			
Instrument	Instruments used to collect measurements			
QA	Exceeded or did not exceed methods acceptable standard deviation			
Topic_begin_date				
Topic_end_date				
Site_ID			x	[WS]_site_ID_table

[WS]_Site_ID_table

Field name	Description	Type and size	Key(s)	Relationship
Site_ID			x	
Site_Descriptor				
Longitude				
Latitude				
GIS_key				

STEWARDS_Units_table

Field name	Description	Type and size	Key(s)	Relationship
Unit reference	Code for referring to unit table		x	
Units of measure	Common abbreviation for units of measure			
Unit description	Describe the units of measure			

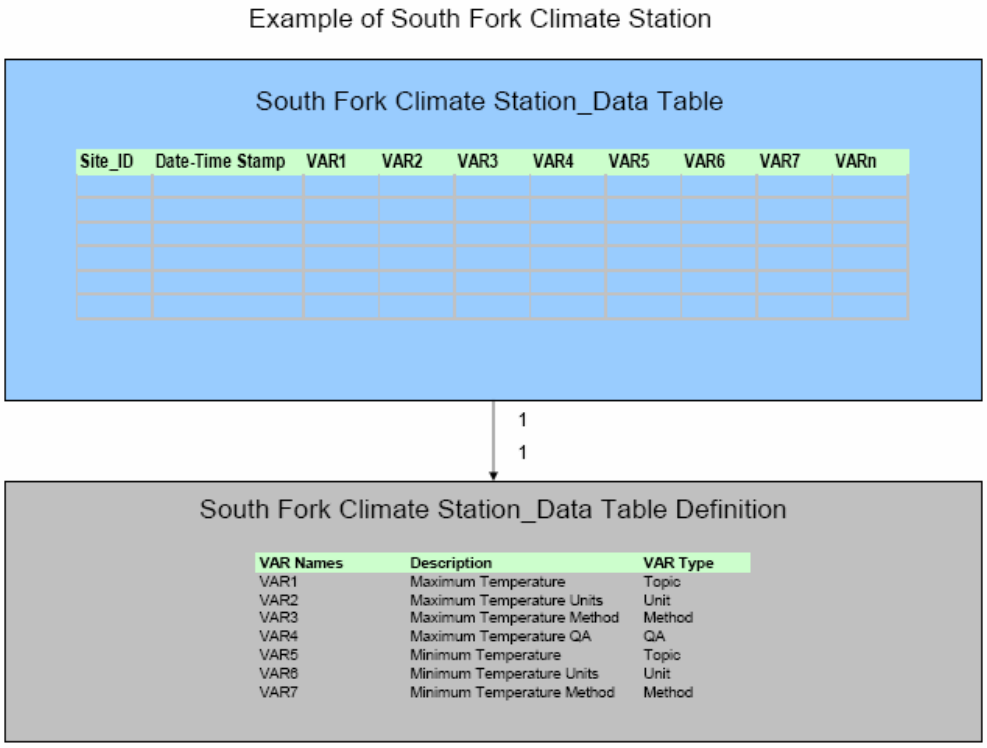


Table F3. Excel style tables and their relationship for a basic data table (Table F1) and a supporting (topic/climate definition) table (Table F2).

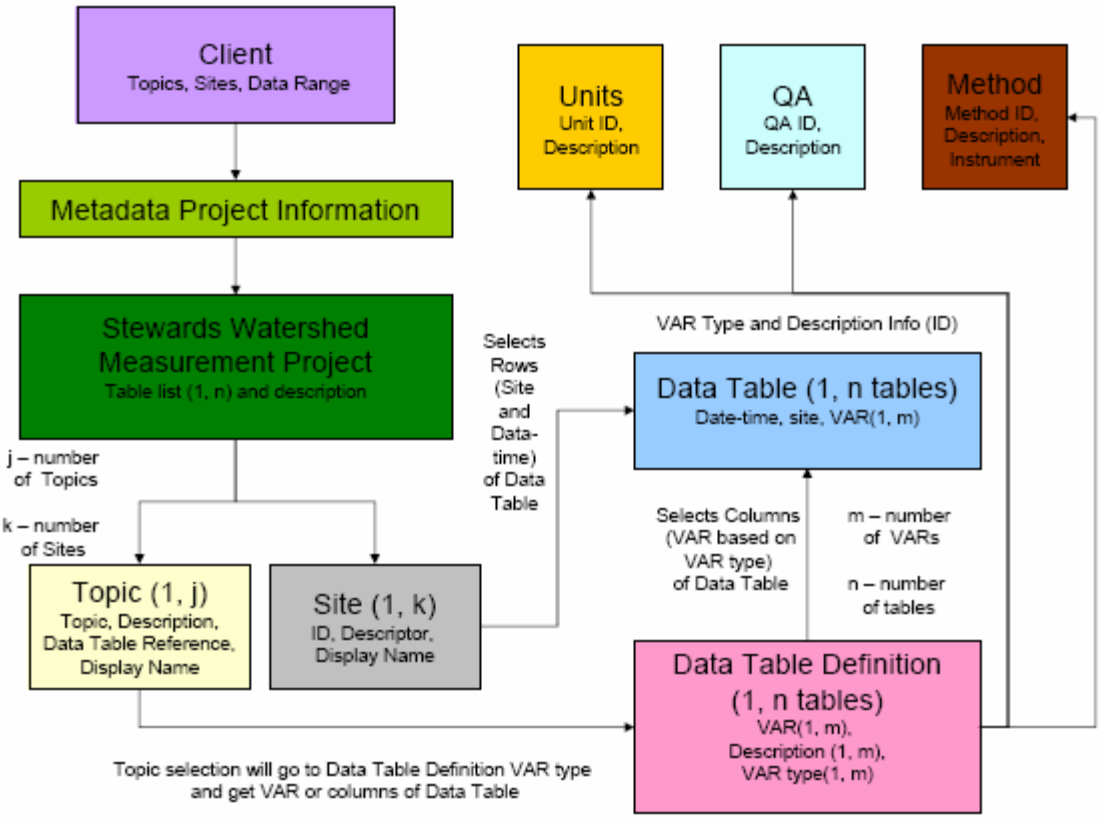


Fig. F4. Generalized data relationships in STEWARDS. The indexes, j,k,m, and n represent the number of topics, sites, variables, and tables, respectively. The lines with arrows show the relationships between tables.

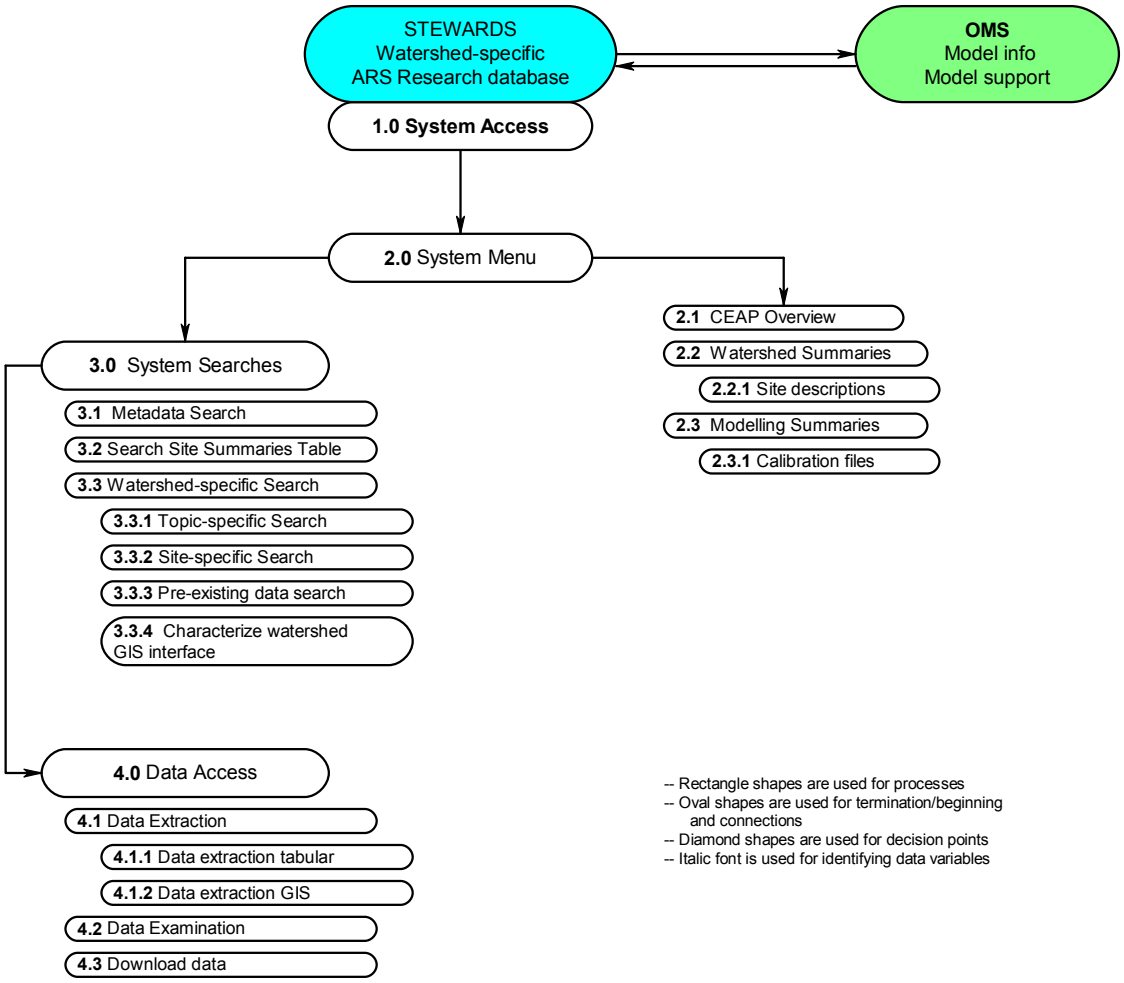


Fig. F5 The application system components of STEWARDS. The numbers represent different application /interface system component at different levels.

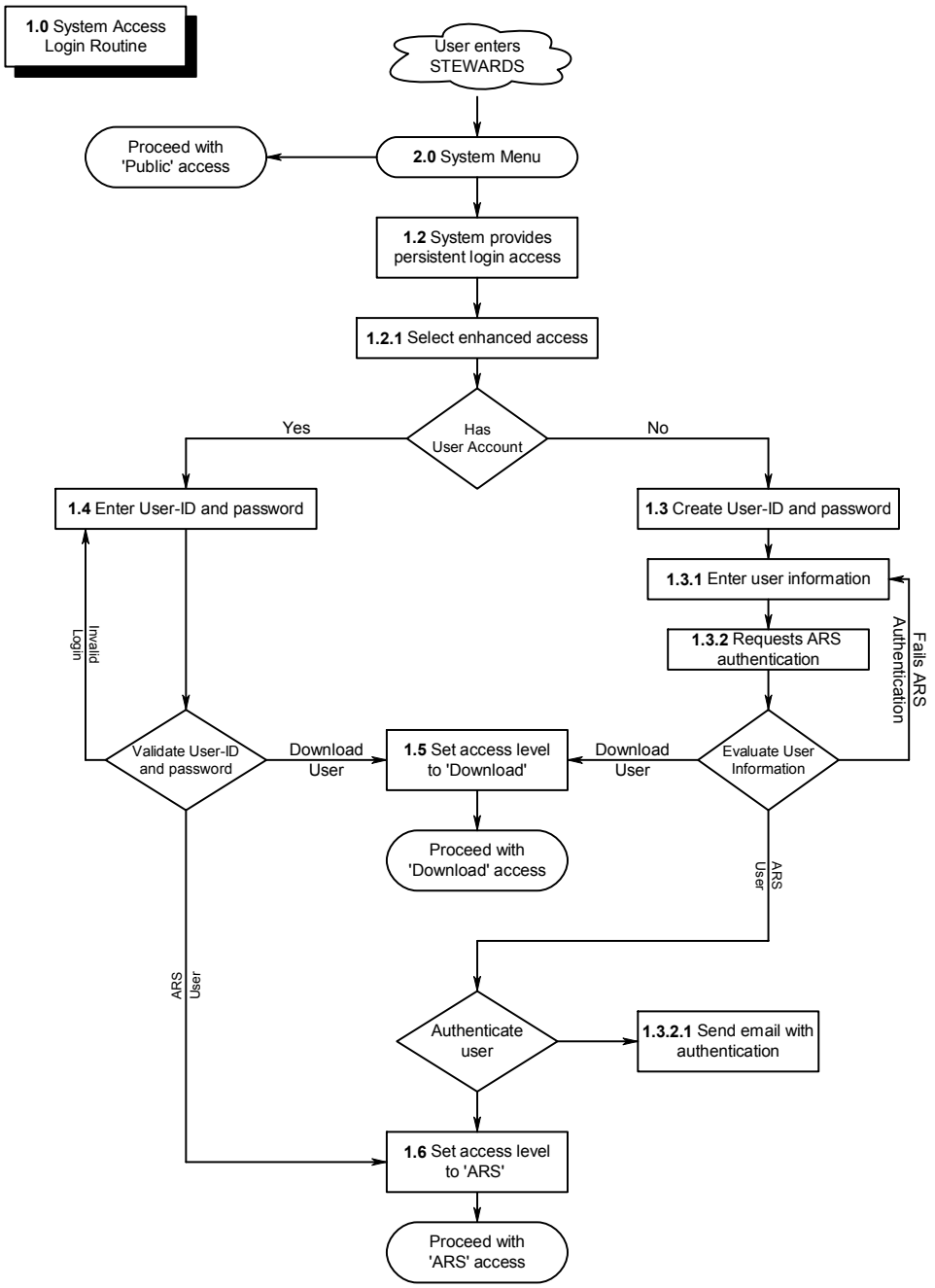


Fig. F6. The 2nd level of access login component of STEWARDS. The first digit of the leading number of each subcomponent represents the first level of the system components and the second number represents the second level and so forth.

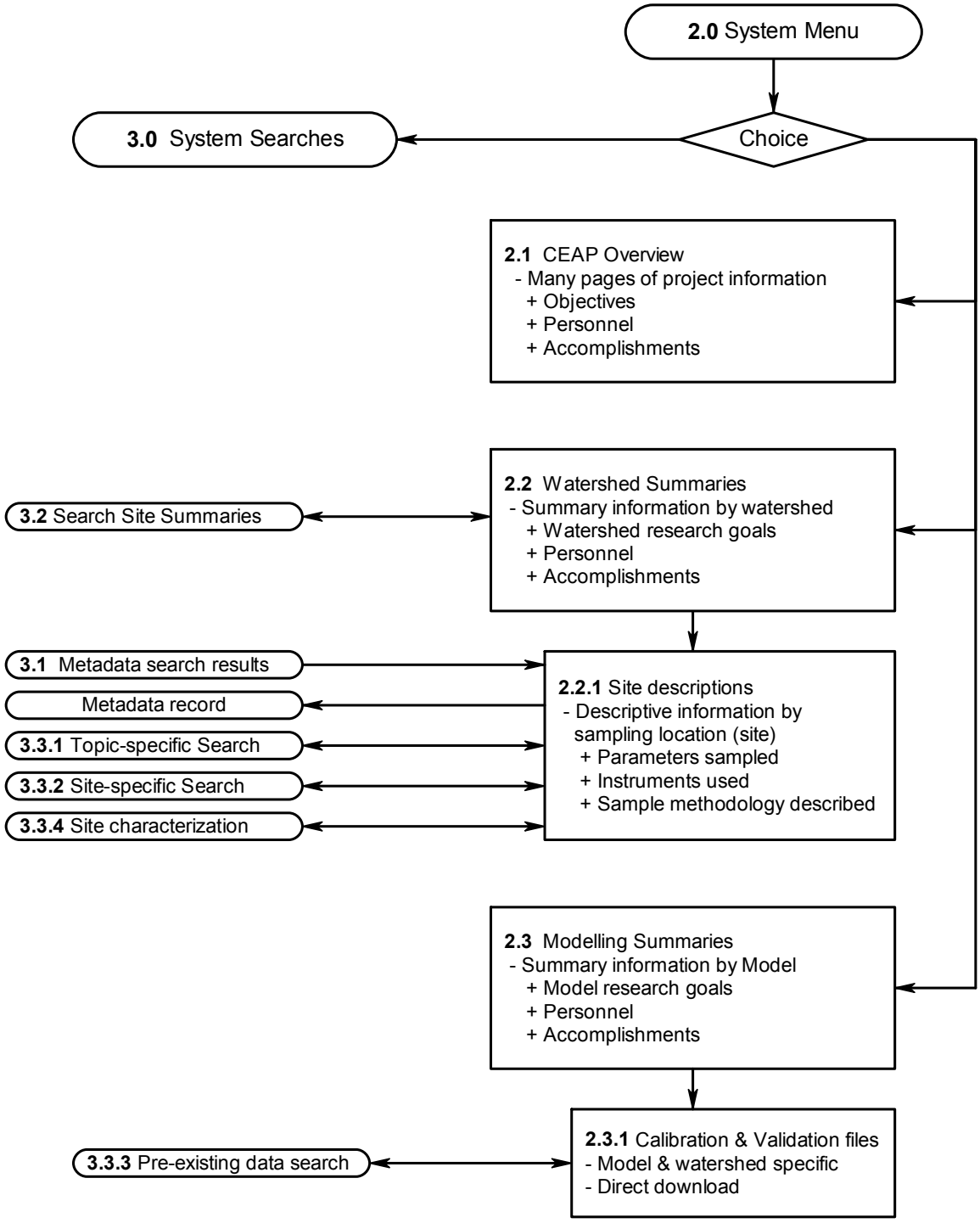


Fig. F7. System/Summary system component of STEWARDS.

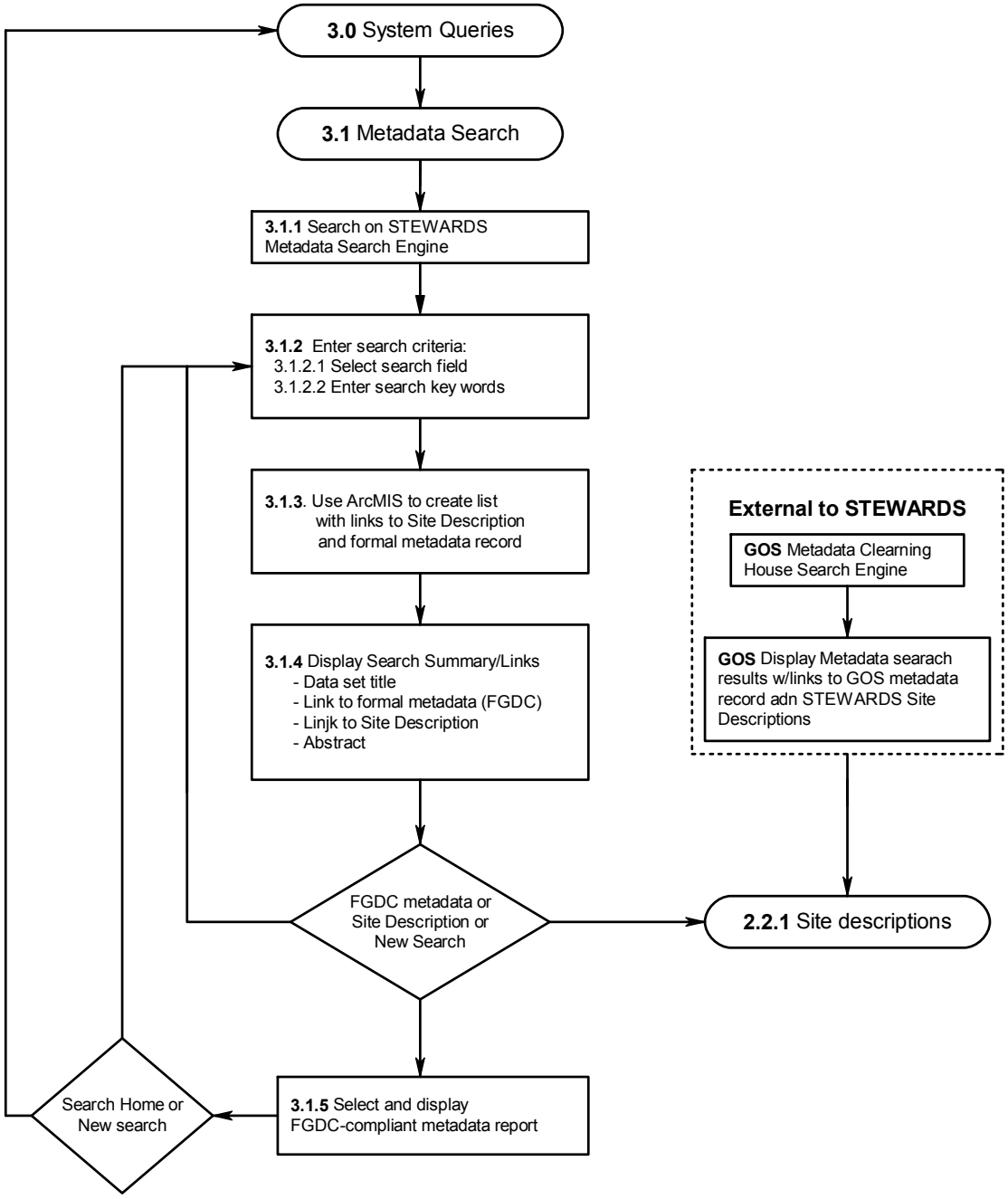


Fig. F8. Metadata search component of STEWARDS.

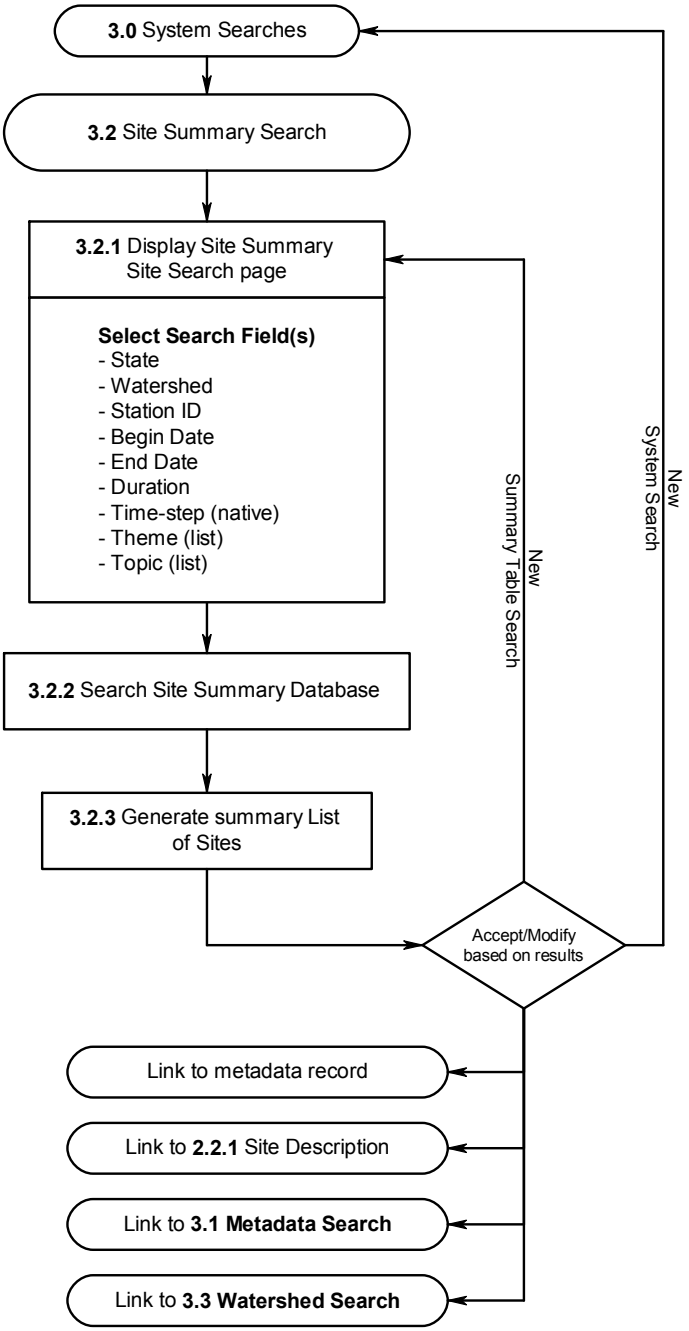


Fig. F9. Site summary search component of STEWARDS.

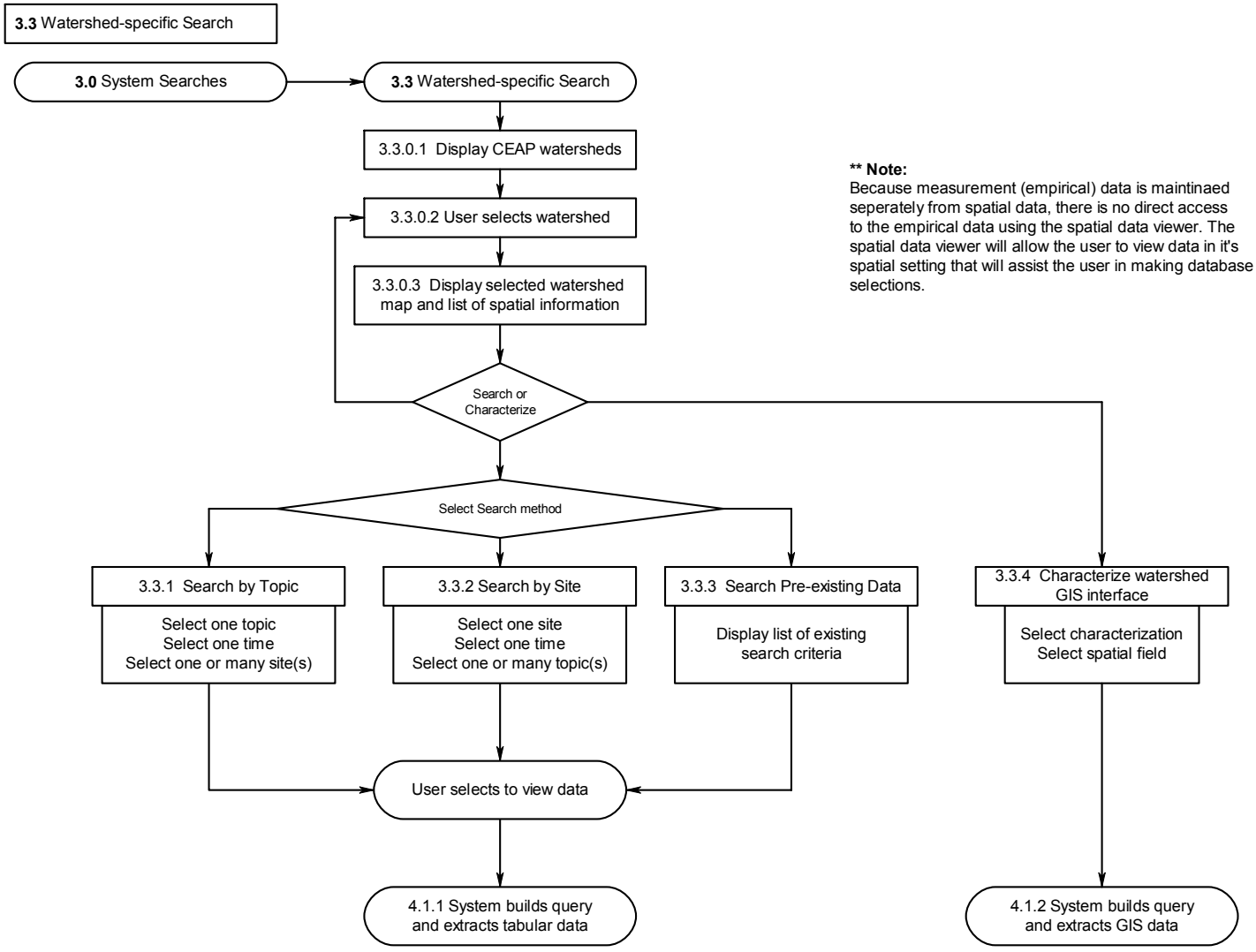


Fig. F10. Watershed specific search component of STEWARDS.

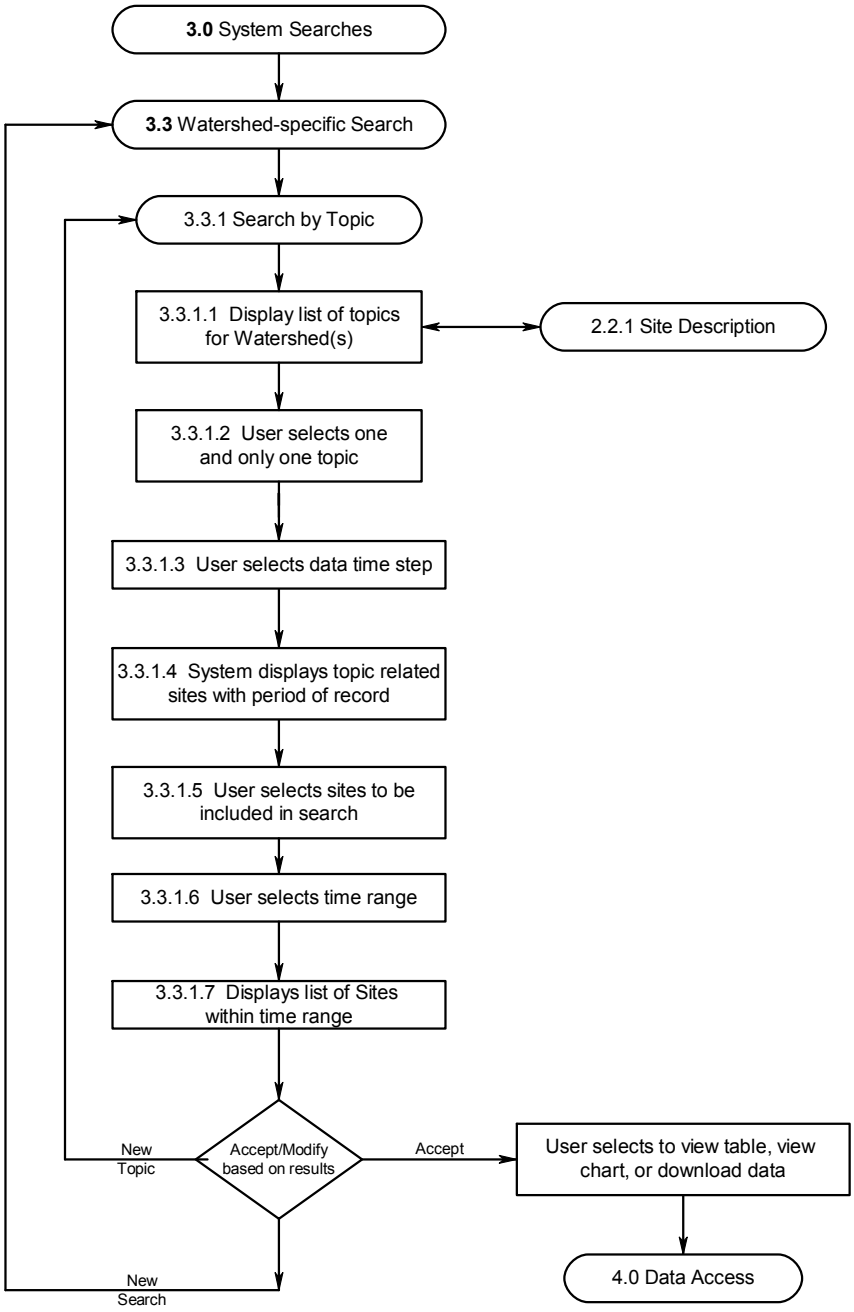


Fig. F 11. Topic search component of STEWARDS.

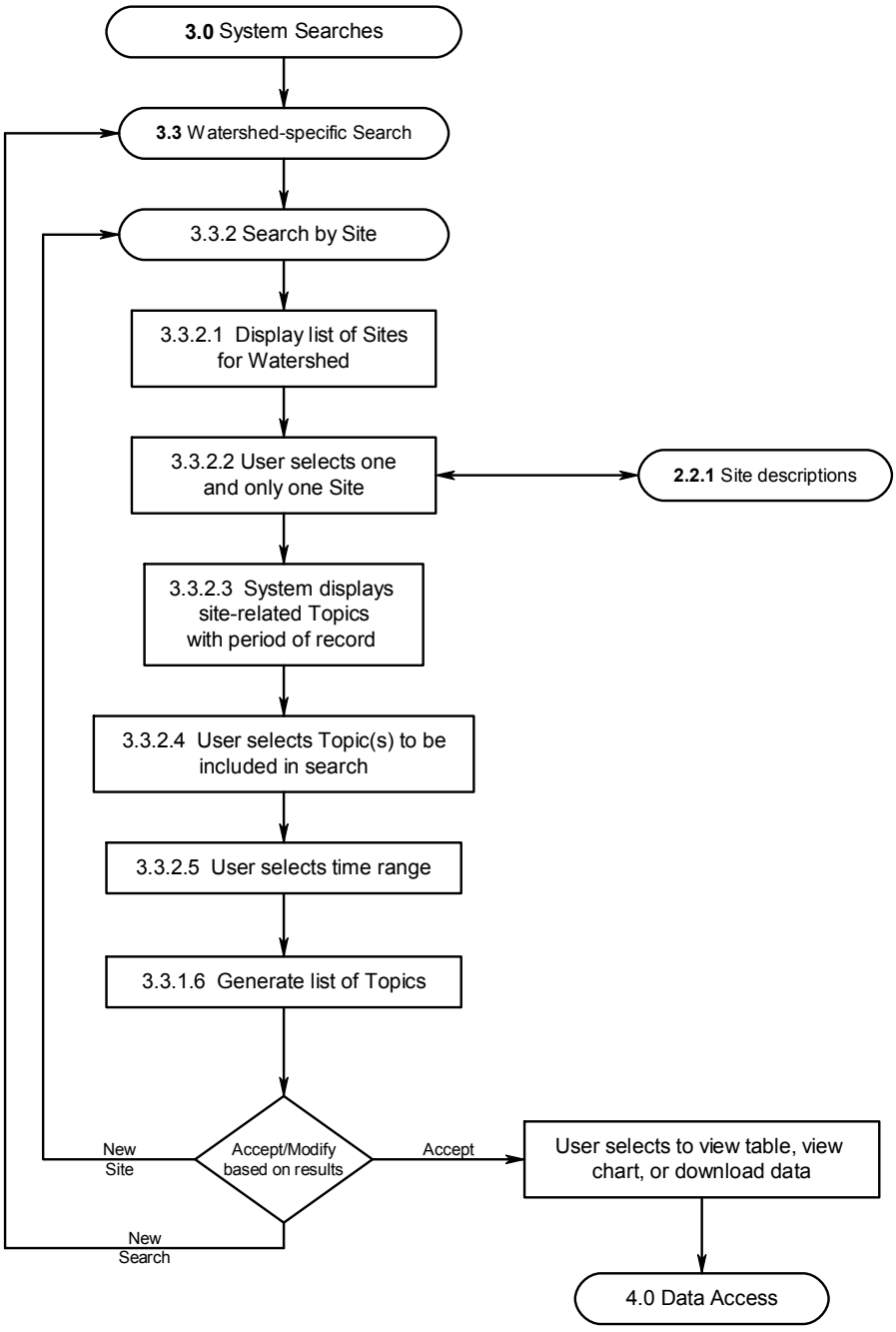


Fig. F 12. Site specific search component of STEWARDS.

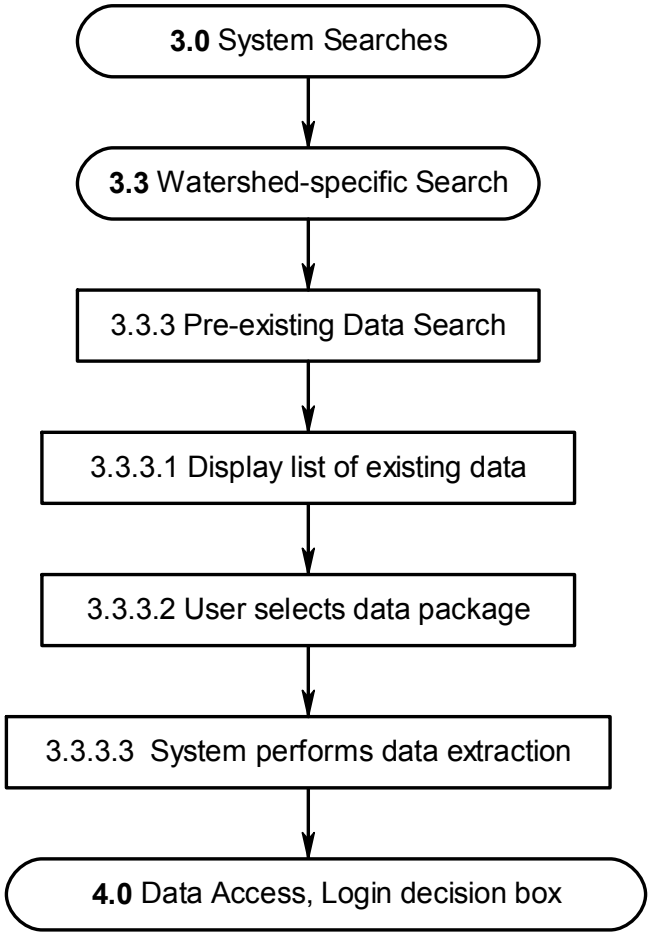


Fig. F 13. Pre-existing search component of STEWARDS.

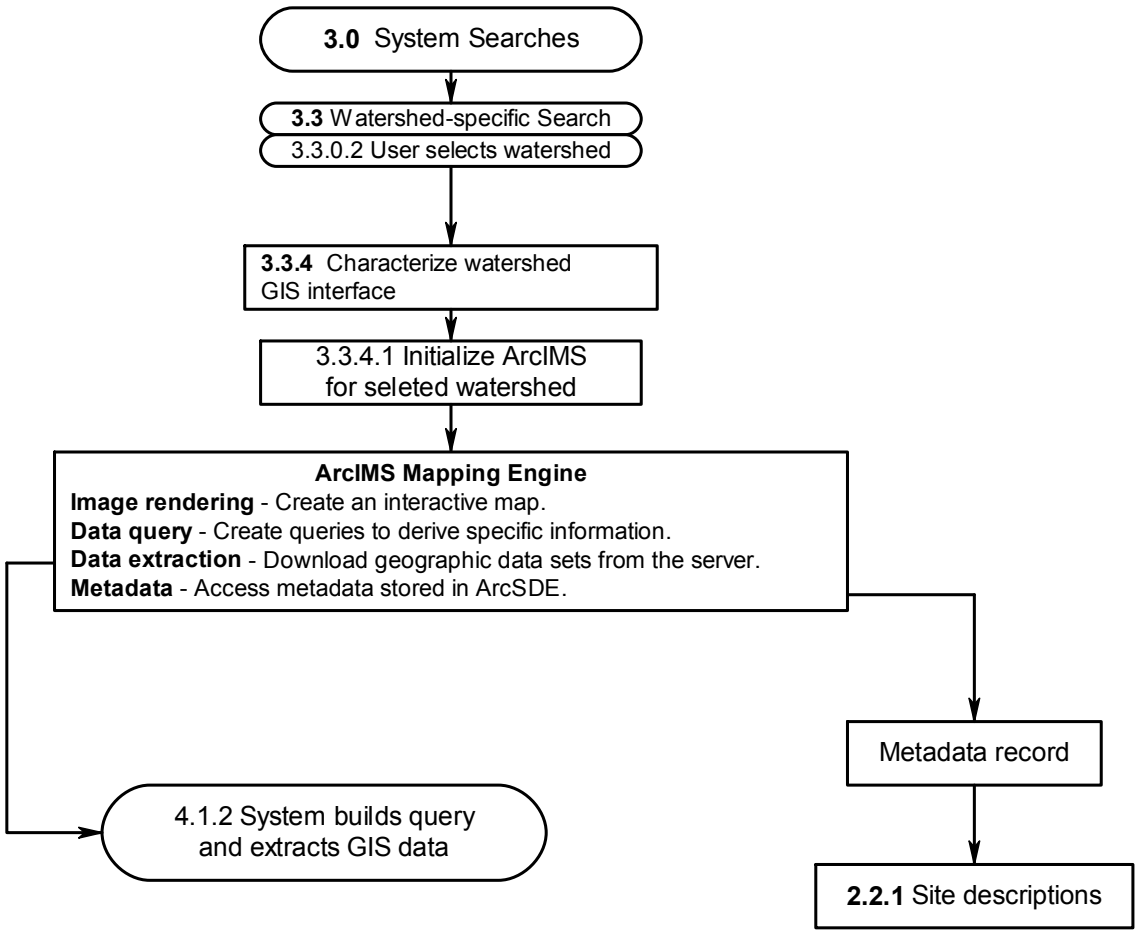


Fig. F14. Characterizing watershed GIS component of STEWARDS.

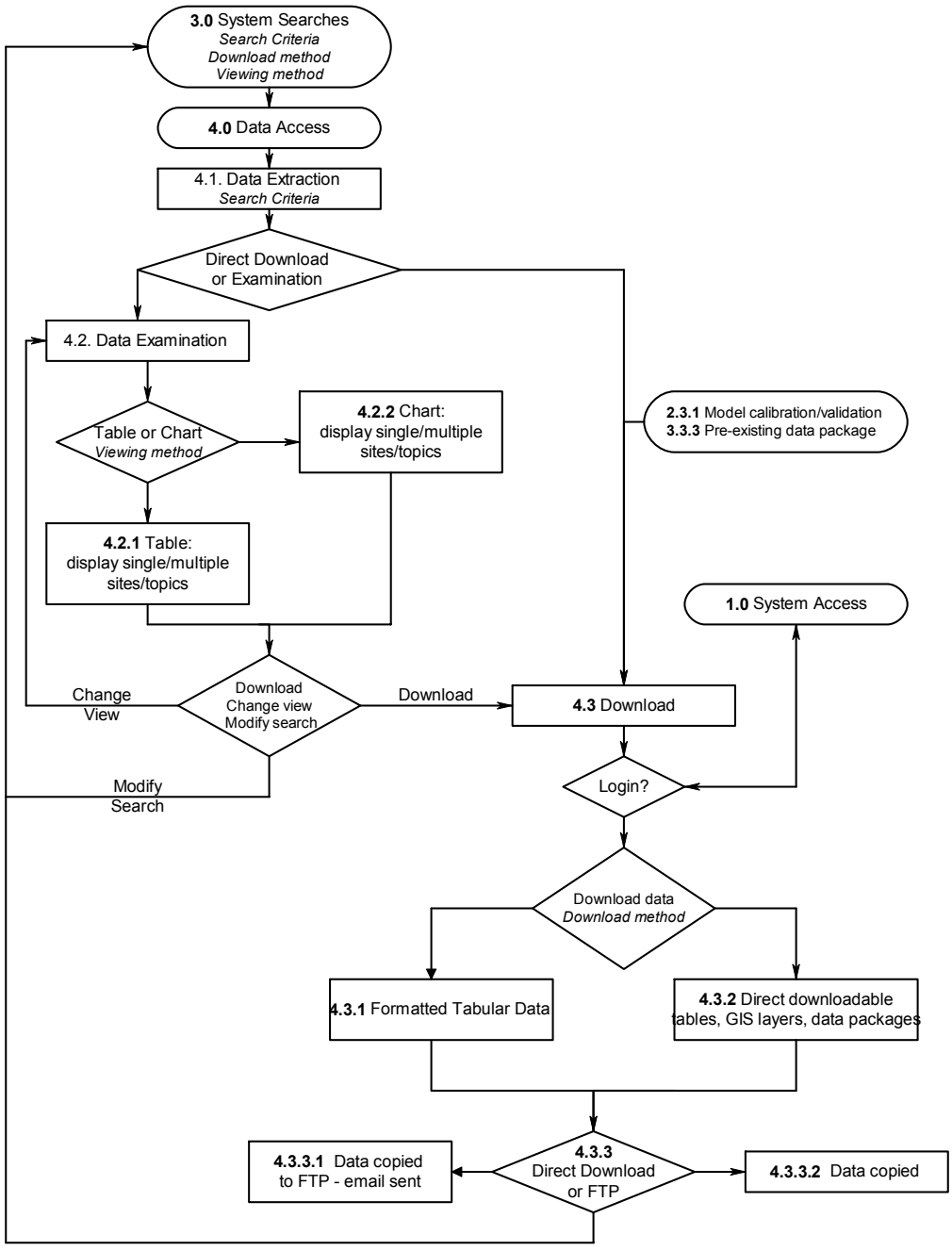


Fig. F15. Data access/visualization component of STEWARDS.

Appendix G: Linkage of STEWARDS to Other USDA/ARS Data/Modeling Systems

As an ARS watershed data system to support SWAT and AnnAGNPS modeling in CEAP, ultimately the STEWARDS data system could be integrated with other similar ARS data/modeling systems such as OMS (Object Modeling System), iFARM (Integrated Farm And Ranch Management), and other general models (Fig. G1).

The **OMS** is a Java-based modeling framework that facilitates simulation model development, evaluation, and deployment. In general, the **OMS** consists of: 1) a library of science, control and database modules; 2) a means to assemble selected modules into a modeling package customized to the problem; and 3) automatic generation of a user-friendly interface. The framework employs the latest Java-based software technology for all the components, and is supported by data dictionary, data retrieval, GIS, graphical visualization, and statistical analysis utility modules.

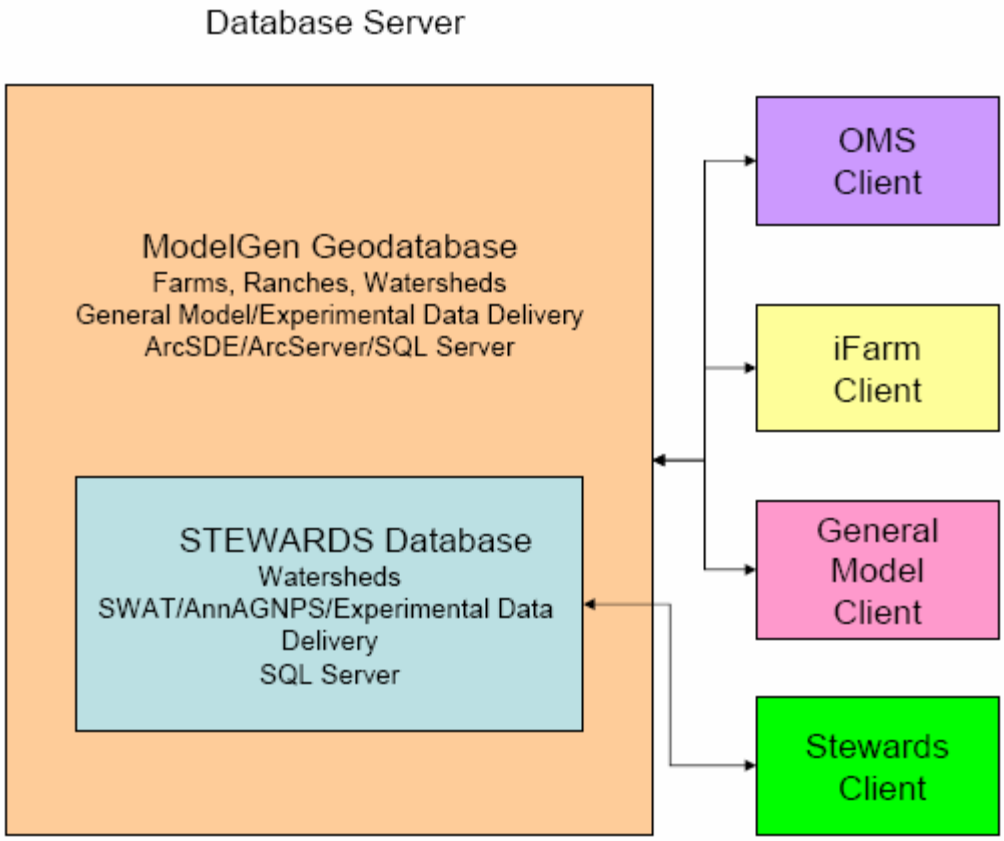


Fig. G1 Linkage of STEWARDS to Other USDA/ARS Data/Modeling Systems.

Appendix H: Task Flow of STEWARDS

The task flow diagram (Fig. H1) of STEWARDS shows that the task components in the development process. The prototype has been implemented but is still under tuning. It should be completed by this summer.

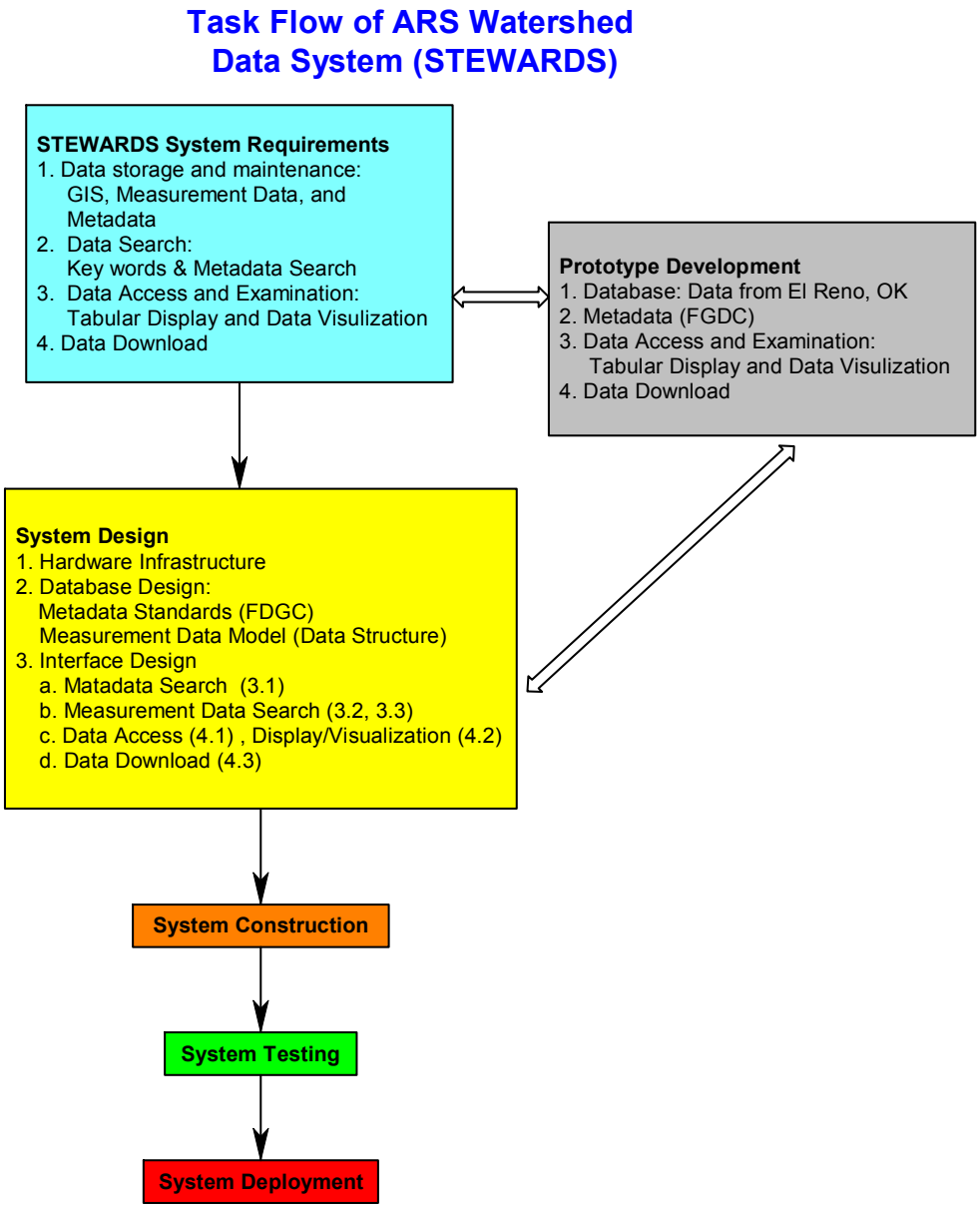


Fig. H1. Task flow of STEWARDS Data System. The number in parentheses represent the system component number as shown in Fig. F15 in Appendix F.

Appendix I. Data Structure for Two Types of Time Series Data in STEWARDS

Type 1: Time series data with columns for different data collected at the same time. Example is flat table of weather data. Fig. I1 shows a generalized data structure for this type of data.
Type 2: Time series data, usually more sparse than the above, with an id field that explains what the parameter field contains. Example is water quality data. Fig. I2 show a generalized data structure for this type of data.

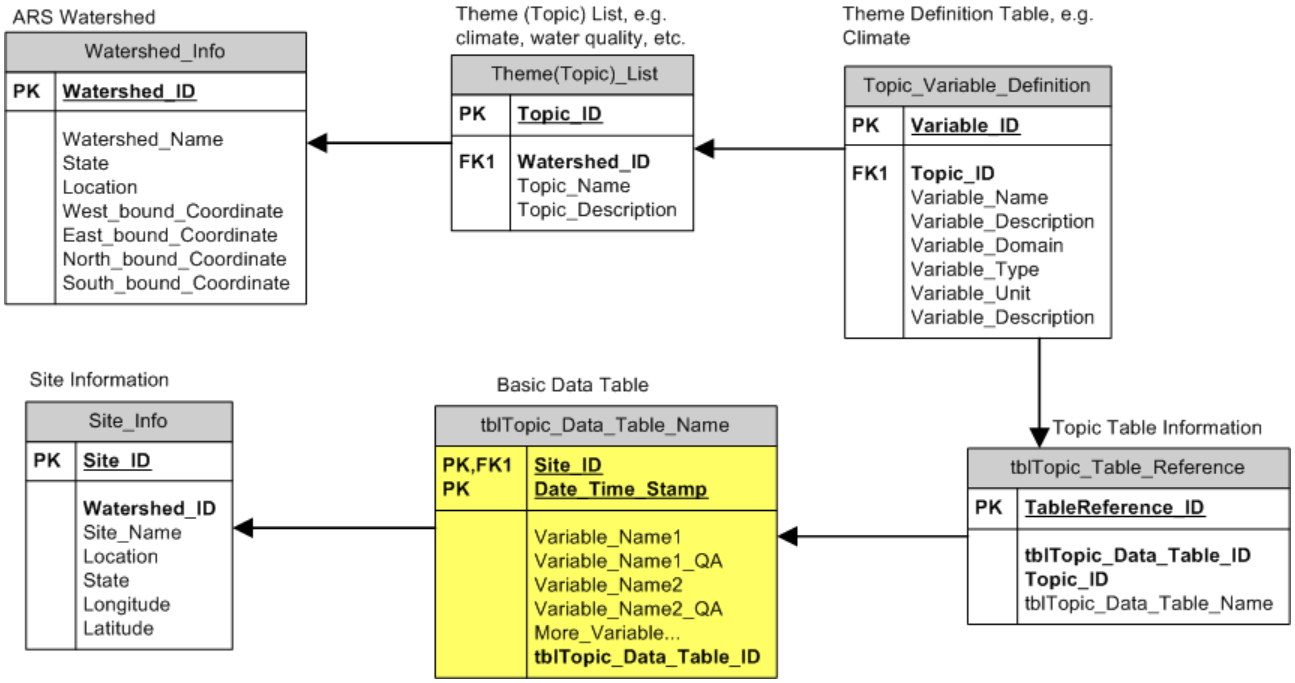


Fig. I1 Data structure for time series data (Type 1 data) with columns for different data collected at the same time.

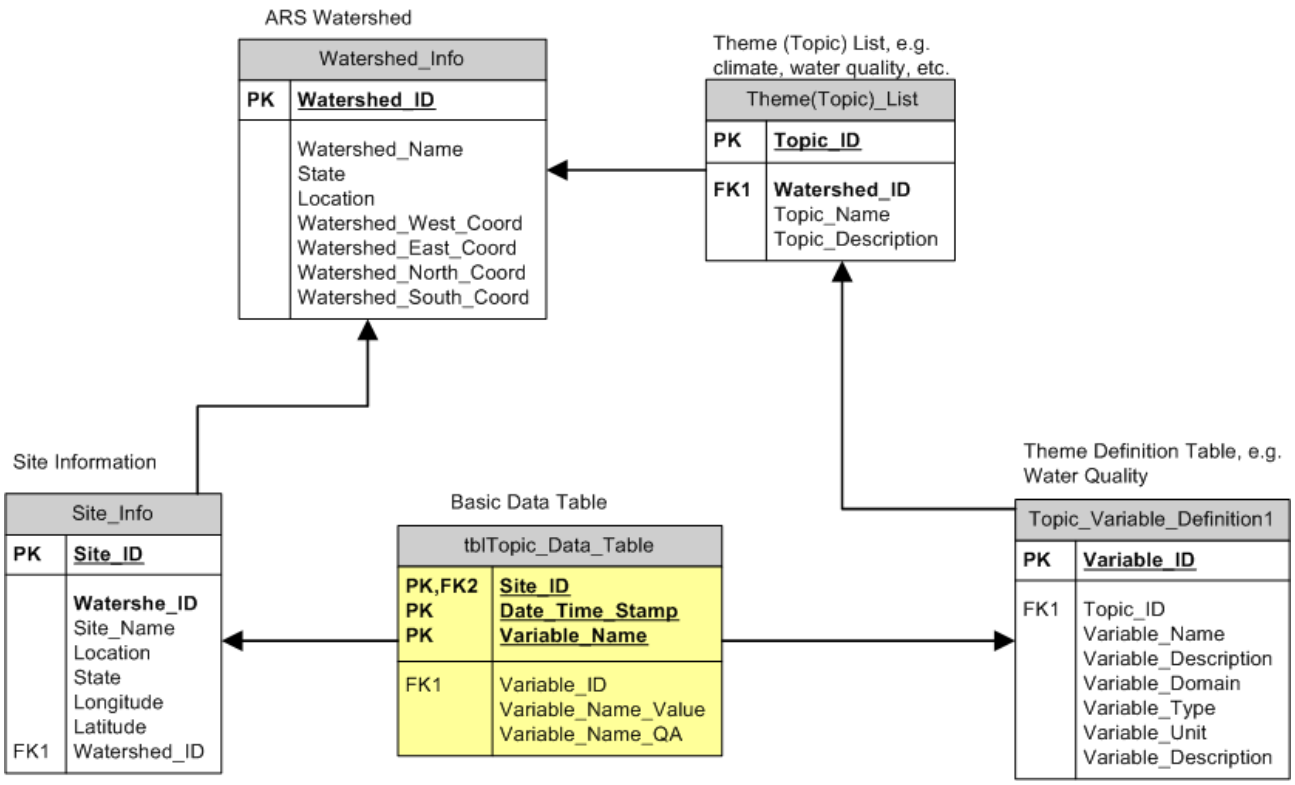


Fig. I2. Data structure for time series data (Type 2 data), usually more sparse than those shown in Fig. I1.