Dynamic Update Techniques for Online Maps and Attributes Data

Thanh H. Pham
Nova Southeastern University, hlua@yahoo.com

Dynamic Update Techniques for Online Maps and Attributes Data

by

Thanh H. Pham

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Graduate School of Computer and Information Sciences
Nova Southeastern University

2001

We hereby certify that this dissertation, submitted by Thanh H. Pham, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Sumitra Mukherjee, Ph.D. — 10-3-2001
Chairperson of Dissertation Committee — Date

Marlyn Kemper Littman, Ph.D. — 10/15/01
Dissertation Committee Member — Date

Emil Horvath, Ph.D. — 6-7-2001
Dissertation Committee Member — Date

Approved:

Edward Lieblein, Ph.D. — 10-15-01
Dean, School of Computer and Information Sciences — Date

Graduate School of Computer and Information Sciences
Nova Southeastern University

2001

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Dynamic Update Techniques for Online Maps and Attributes Data

by
Thanh H. Pham

April 2001

Online databases containing geographic and related tabular data for maps and attributes often require continuous updates from widely distributed sources afield. For some applications, these data are dynamic, and thus are of little value if they do not reflect the latest information or changes. A status map that depicts graphically temporal data affecting accountability is an example of this type of data. How can accommodations be made collectively for the perpetual data updates in the database and the need to deliver online information in real time without making concessions? The goal of the dissertation was to analyze and evaluate techniques and technology for data collection and storage, online data delivery, and real-time upload. The result of this analysis culminated in the design and prototype of a system that allowed real-time delivery of up-to-date maps and attributes information. A literature review revealed that an ample amount of research material existed on the theory and practice of developing dynamic update techniques. Despite that fact, no research literature was available that specifically dealt with dynamic update techniques that provide for real-time delivery of up-to-date maps while allowing online update of attributes information. This dissertation was the first attempt at providing research material in this important area. The procedure consisted of five major steps encompassing a number of small steps, and culminated in the development of a prototype. The steps included gathering data collection and storage information, investigating technological advances in data delivery and access, studying dynamic update techniques, assessing the feasibility of an implementation solution, and developing a prototype. The results revealed that the dynamic update technique as implemented in the prototype met the need for timely delivery of accountability, geospatial, and metadata information within an infrastructure.

Acknowledgements

This research was partially supported by the U.S. Department of Agriculture - Natural Resources Conservation Service - National Cartography and Geospatial Center.

Special thanks to Dr. Sumitra Mukherjee, my dissertation advisor, for his guidance, encouragement, and thoughtfulness. Many thanks to Dr. Marlyn Kemper Littman and Dr. Emil Horvath for their service as my Dissertation Committee members. I will always be indebted to Dr. Horvath, Mr. Paul Fukuhara, and many others at the National Cartography and Geospatial Center for their help, patience, and understanding. I would also like to express my appreciation for the assistance from the collaborative project with the Department of Urban Studies and Planning at Massachusetts Institute of Technology.

To my parents, wife, and children for their unconditional love, and support.

Chapters

List of Figures

Figures

Chapter 1

Introduction

Problem Statement and Goal

Online databases containing geographic and related tabular data for maps and attributes often require continuous updates from widely distributed sources afield. For some applications, these data are dynamic, and thus are of little value if they do not reflect the latest information or changes. A status map that depicts graphically temporal data affecting accountability is an example of this type of data. How can accommodations be made collectively for the perpetual data updates in the database and the need to deliver online information in real time without making concessions? As an example, the National Cartography and Geospatial Center (NCGC) in Fort Worth, Texas is a production center of the USDA's Natural Resources Conservation Service (NRCS). NCGC is responsible for the technical leadership, coordination, archive, and eventual dissemination of eleven-or-fourteen–digit hydrologic unit (HU) data. Teams of experts statewide are accountable for the compiling and digitizing of these HU data. A report on the status of this project would involve almost a hundred telephone calls, faxes, and relentless pursuits. The report would probably take a month, and would nonetheless be little more than a guess in the end. The desired objective was for an online reporting process, which is accessible by all participants. If such a process existed, the above report should take only a couple of minutes. Orland, Wu, and Chavan (1997) discuss the suitability of online systems for supporting communities and agencies in natural resources planning. Litwin, Mark, and Roussopoulos (1990) emphasize the need for shared access. Executive Order No. 12906 (1994) signed by President Clinton, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," mandates that agencies adopt methods for documenting geospatial data previously collected or produced, and make that data documentation electronically accessible to the National Geospatial Data Clearinghouse network. Faundeen and Zanter (1997) describe the U.S. Geological Survey (USGS) where metadata representing millions of cartographic, topographic, and remotely sensed image data products are updated daily and are made accessible to the public. Using internal maintenance and external public databases, USGS provides its personnel and interested public users with accurate and timely information. Users can query and retain most of this information through online reports.

The more progressive teams within NRCS, of which NCGC is a part, use project management software like Microsoft Project. These teams post detailed project plan charts, such as Gantt and PERT (Program Evaluation and Review Technique), to Web pages. Despite how plausible this procedure may seem, it is far from being dynamic. NRCS is a federal agency whose mission is to provide leadership in a partnership program to help people conserve, improve, and sustain our natural resources and environment (USDA - Natural Resources Conservation Service [NRCS], 1994b). In 1997, NRCS's vision of information technology calls for user-friendly information technology tools, capability, and supporting infrastructure that provide real-time access to data and information in the field (USDA - Service Center [SC], 1997). NRCS's strategic plan is to get the right information, in the right form, to the right people, at the right time (USDA - NRCS, 1996). The Geospatial Database Branch (GDB) of NCGC is a national data clearinghouse. The number one goal specified in GDB's 1997 business plan calls for providing effective and efficient online access to tabular and spatial data directly from the database over the Web (Horvath, 1996). In addition, the USDA's directives regarding departmental – administrative convergence, moratoria, and current technology developments all lead us to believe that cheaper, faster, and more relevant (i.e., dynamic, up-to-date) is the direction that information-technology communities must take to serve their business clients more advantageously.

The goal of the dissertation was to analyze and evaluate techniques and technology for data collection and storage, online data delivery, and real-time upload. The result of this analysis culminated in the design and prototype of a system that allowed real-time delivery of up-to-date maps and attributes information. There were four specific design goals for this system: Web delivery; dynamic, mostly unattended upload; continuity and robustness; and implementation plan for NCGC. Web delivery fulfilled the requirements for a consistent, ubiquitous, and platform-independent interface. Benefits of Web technologies were demonstrated by Balasubramanian, Bashian, and Porcher (1997) in a case study on how a large-scale hypermedia authoring and publishing system was designed to satisfy authoring, management, and delivery needs; and by Lush (1996) in the development of a Web-deliverable electronic textbook. Up-to-date information in the field must be available quickly for viewing, analyzing, or ad hoc studies regardless of where the user is. Therefore, data would upload dynamically to a central database, and to any remote location, without additional work from the field. A similar endeavor was demonstrated by Caputi (1998) in the development of a software solution for real-time digital signal processing (DSP) where real-time communication support, such as uploading data for displays and file output, is transparent to the user. The procedure to update information in the field was independent of the individuals involved (i.e., a change of personnel does not affect the integrity of entered data), and accommodated dissimilar users sufficiently. One of the research priorities presented by the University Consortium for Geographic Information Science [UCGIS] (1996) was the representational robustness needed to integrate diverse data across a wide range of applications and disciplines. The implementation plan emphasized NCGC, although it may be relevant in other environments. The NRCS Clearinghouse business plan (Horvath, 1996) maps out a strategy for online and on-demand delivery of the agency's data assets in the public trust. The requirement is for reliable and efficient access through Web browsers, made possible by storing all data types in a relational database management system (RDBMS), and the implementation of spatially enabling technology. System reliability and performance were important factors in the implementation plan that made assumptions about NCGC future technology infrastructure including client access, bandwidth, distribution method, and support.

Stiles, Tewari, and Mehta (1997) describe human–computer interaction techniques for using VRML (Virtual Reality Modeling Language) 2.0 models in an immersive virtual environment. They mention that VRML offers an excellent choice because the method for dynamic update of 3D scenes over the Internet is built into the standard. Acquah (1991) describes geometric modeling and interactive graphics techniques using the Imagys language, which supports dynamic update and direct manipulation of data objects. Similarly, dynamic update techniques were used in the proposed prototype development, which met the design goals described in the previous paragraph. The analysis and design of the prototype were based on principles from Whitten, Bentley, and Barlow (1994). Coad and Yourdon (1990) provided the conceptual foundation for the Object-Oriented (OO) analysis and design methodology. Guidelines for the OO software development were incorporated from Rumbaugh, Blaha, Premerlani, Eddy, and Lorensen (1991). Practical tips and techniques from Horstmann (1997) were applied during the programming phase. Elmasri and Navathe (1994) provided the basis for the design and implementation of the database system. Trends and techniques considerations in data communications and network came from Stallings & Slyke (1998). Policy and procedures pertinent to the development of systems for online delivery of geospatial information and related attributes were consulted from USDA - NRCS (1995), USDA - NRCS (1996), USDA - SC (1997), and USDA - SC (1998).

Relevance and Significance

NRCS has a large repository of geospatial, tabular, and document data. These data receive continuous updates from various sources. NRCS wants to serve these data to employees and to the public effectively and timely.

In 1996, NCGC became the official data clearinghouse for NRCS (Folsche, 1996). NCGC archives and distributes geospatial data, such as MUIR (Map Unit Interpretation Record), SSURGO (Soil Survey Geographic), STATSGO (State Soil Survey Geographic), NRI (Natural Resources Inventory), SSL (National Soil Survey Laboratory Characterization), and DOQ (Digital Orthophotography Quadrangles). Various NCGC staffs and individuals participate in the acquisition, validation, and certification of these data that come from sources nationwide. NCGC replenishes and updates its databases continually. Some data, such as SSURGO, are more dynamic and require updates more often than the others. NRCS generates some certified geographic soil data, such as SSURGO and STATSGO, at the county and state level. It shares the cost in the production of others, such as DOQ, with USGS and FSA (Farm Service Agency). Total data anticipated by the end of the year 2001 is 25 terabytes. NCGC is currently involved with integrating and reformatting data from various databases. It will eventually be involved with gathering different data from other sources to create integrated data sets at the county level. GIS (Geographic Information System) tools, such as ArcInfo, are used during verification to eliminate erroneous data. Only SSURGO and STATSGO data are available for downloading by FTP (File Transfer Protocol). NCGC currently disseminates most of its data on CD-ROMs and plans to be on line with all NRCS data (Folsche, 1996). In addition, to meet the challenge put forth in "America's Private Land, A Geography of Hope" (USDA - SC, 1996), NRCS must provide field conservationists with real-time access to data and information (USDA - SC, 1997).

Besides geospatial data, the NCGC data repository consists of tabular data marts holding critical business information. It is the dynamic nature of daily work that these data undergo persistent updates. NCGC has developed business applications using 4GLs (Fourth Generation Languages) as an interface to these databases. The goals have been to help managers, data stewards, and employees with critical or strategic decision making, accountability (or progress reports), and inventory requests. For years, these goals, among which accountability and progress are particularly important, have met with some success. The center director has made it a priority to improve and replace these archaic information systems to meet those goals effectively.

NCGC also maintains hundreds of conservation standards, practices, and job sheets on line. These documents, which are available for FTP or Internet download in PDF (Portable Document Format) or Word-for-Windows format, habitually undergo revisions (USDA - NRCS, 1994a).

The desire for dynamic, up-to-date information is apparent for all three kinds of data (i.e., geospatial, tabular, and documents). Croft and Savino (Croft & Savino, 1998), addressing the importance of providing efficient and effective methods, emphasize this fact in a recent paper for the storage and retrieval of text and document data. The advantage of information is so strategically important that the ability to deliver online information updates expeditiously is more essential than ever.

The USDA - SC (1998) found that integrating GIS technology into service center business operations is crucial to provide timely program delivery and accurate geospatial information. Such a system would provide common access to consistent data and geospatial business processes among service center locations and service center agencies. It would eliminate spatially inaccurate and expensive paper-based maps and information, eliminate duplicate sets of information and processes, and provide easy online access to geospatial information.

Recognizing the problem of organizing the large number of documents in online digital libraries, Orendorf and Kacmar (1996) proposed a spatial method of structuring digital libraries and their content in which users navigate geographically to locate and access information. They developed the Spatial Document Locator System (SDLS) and found it to be well suited to the task of providing access to the multitude of geographically based (spatial) documents, such as GIS data sets.

The publications group at MIT (Massachusetts Institute of Technology) had a problem of overlapping and redundant documents information. Jones (1994) analyzed the existing documentation set in light of the extended computing environment at MIT, then proceeded with the transition to an online hypertext documentation system.

Merrill Lynch wanted to provide instantaneous access to current financial information to financial consultants, other professionals across the corporation, and the public through the Internet. Balasubramanian, Bashian, and Porcher (1997) designed and implemented a large-scale hypermedia authoring and publishing system using document management and Web technologies. The system satisfied all Merrill Lynch requirements including distributed environment, consistent user interface, reduced maintenance, access control, concurrency control, document management, and full-text and attribute-based information retrieval.

Adam et al. (1996) presented a study on strategic directions in electronic commerce (EC) and digital libraries (DL). Key challenges suggested in the study include facilitating the update of existing content and the management of multiple versions of objects, providing methods to capture continuous media in real time, and providing uniform user interfaces that take advantage of specialized data types such as maps.

At the University of California, Berkeley, Wilensky (1996) suggested the technical solutions for a work-centered digital information system that met users requirements for effective information extraction, better accessing of information, and improved interaction with repositories. The focus was on large and diverse environmental information, which includes millions of pages of technical reports; aerial and ground photography; USGS topographic, land use, and other special-purpose maps.

Evans (1997) built a prototype Web browser for digital orthophotos as a part of his dissertation. Orthophotos are tens to hundreds of megabytes in size. They require generous amounts of storage space, powerful tools to extract and rescale image portions for particular uses, and GIS expertise to position the image correctly behind a map. The prototype facilitates the finding and retrieving of only the image portions needed, compressing them for efficient use of a limited bandwidth, and providing the header files as needed to integrate the image portion into GIS maps.

Harder (1998) suggested in his book that the convergence of GIS and the World Wide Web (WWW) has changed mapmaking forever. Detailed maps can now be generated on demand from huge databases of spatial information and transmitted instantly across the globe. By getting GIS on line, the Internet makes possible the accessibility of geographic information and the power of GIS applications to solve problems.

On the topic of computer-assisted thematic cartography, Tyner (1992) described in his book the advantages of computers in cartography, which are speed, cost, and flexibility. Many time-consuming and tedious chores can be handled by computer. Calculations can be made more rapidly and automatically from online data. It is more cost-effective to make a great number of maps on the computer. The use of the computer for mapping allows practical and cost-effective experimentation with different representations of the same data.

Environmental Systems Research Institute [ESRI] (1998) presented a geographic data explorer designed to view and query geographic data stored on the computer or on the Web. This tool solves one of the most pressing concerns regarding geographic data and Web applications—how to distribute data openly and freely.

Rowson (1998) stated the possibility for the delivery of electronic design automation (EDA) tools over the Internet, especially in the area of incremental library updates. Strong Web features, such as interoperability and ubiquitousness, will increase reliance on the Internet for design documents.

Orland, Wu, and Chavan (1997) believed the Internet provides great opportunities to advance existing information generation, storage, dissemination, and retrieval techniques for environmental management. The Internet is also considered as an interactive medium for providing timely information services, which won't be possible via traditional communication services such as TV, telephone or newspaper. Moreover, the Internet provides a highly dynamic mechanism for customizing information to meet the ever-changing needs of different types of users.

Entlich et al. (1997) suggested a model for the conversion of large text and graphics collections to an electronic format that will serve the document delivery needs of scholars in a distributed networking environment. The model provided a realistic environment in which to investigate the technological problems associated with full-text retrieval and delivery.

Bezanson (1995) introduced an electronic performance support system (PSS) to allow workers to control their own learning by giving them the ability to retrieve information at the workplace when they need it. A PSS, which is embedded within the overall system or environment that the worker is using, provides just-in-time training, information, and help functions on a system or product.

Kirstein and Montasser-Kohsari (1996) piloted a project to provide online access to a large collection of electronic journals. They experimented with various forms of data representation and with indexing data and retrieval mechanisms. Using both SGML (Standard Generalized Markup Language) for authoring and ODA (Office Document Architecture) for distribution, they demonstrated the technical feasibility and utility of electronic access to these documents.

Recently, Schatz and Chen (1999) stated that the WWW has made access to the Internet part of the structure of everyday life. Griffin (1999) observed that millions now regularly use the Web as a primary source of information, and as an inventive medium for communicating and sharing knowledge, enabling new relationships, collaborations, and intellectual communities. Corn (1999) observed that health-related activities depend on vast seas of information, and gave an example of the National Library of Medicine—the world's prime repository of biomedical information. Schatz, Mischo, et al. (1999) suggested the development of mechanisms to provide effective access to full-text physics and engineering journal articles within an Internet environment. Thomas, Alexander, and Guthrie (1999) reasoned that providing access to older scholarly materials—by converting them to digital media and providing full text search capability—would ensure the preservation of these materials and save library space.

Studies in literature demonstrated the needs and benefits associated with the ability to deliver and access information on line. This dissertation fills an acknowledged gap in current delivery technology; it advances techniques to provide online access and real-time update of dynamic spatial data and attributes effectively and efficiently. This investigation is relevant not only to NRCS but also to the public in general.

Barriers and Issues

An industry-wide consensus on how to accomplish quick delivering of online updates has not been reached because of technical, organizational, or administrative reasons. Technologies that are relevant to posting and maintaining online data continue to evolve rapidly. New products are developed and marketed constantly. Those technologies that are technically satisfactory either fail in the marketing stages or confine themselves to a product line. It takes time to separate the trends from the technology. For example, compression technologies, such as MrSID (Multi-resolution Seamless Imagery Database) from Lizard Tech, which is claimed to be the most powerful image compression software (Lizard Tech, Inc., 1999) and considered effective in saving storage space for large image repositories, are still in early development stages and not easily applied in practical situations. Web performance still falls short of expectations in certain areas, such as bandwidth and latency (Rowson, 1998). It takes a fast network to communicate quickly. Slow network connections continue to degrade run-time performance significantly (Geppert, 1998). The integration of mobile computers within existing networks, an NRCS vision of information technology (USDA - SC, 1997), also poses a new set of problems (Acharya & Badrinath, 1996).

An administrative issue exists concerning access control. Implications of access to the files system (i.e., data and system security) and associated misunderstandings compound to create public mistrust of dynamic update techniques. Webcasting techniques, such as Push technology, have had their share of critics for fear of going against the spirit of the Internet-seeking and retrieving information from the sprawling network as it is needed (Rosencrantz, 1997). Another administration issue is financial support. Governments are under pressure from disgruntled taxpayers and face spiraling costs. Research is one of the easiest budgets to attack, because the kinds of legislation that preserve pensions and other entitlements rarely protect it (University Consortium for Geographic Information Science, 1996).

Organizationally, ongoing needs for appropriate technical expertise in the management of information technology (IT) continue. There is the question of how well middle and upper level managers are equipped to meet the management challenges of IT (Fry, 1996). Managers have yet to realize the potential of IT to make a significant contribution to improving client services (Broadbent, Butler, Hansell, & Dampney, 1995). Still lacking are geographic information infrastructures allowing managers to share information across organizational boundaries meaningfully (Evans, 1997); and geospatial data standards for transfer or exchange, metadata, data content, data quality, data models, and data collection (Wortman, 1994).

The volume of geospatial data is also an issue. Online retrieval of huge volumes of text and image data continues to be problematic (Entlich et al., 1997). Although the hardware and software offerings are growing dramatically, they still have limitations and are costly. Many approaches developed and demonstrated by universities and vendors over the past years have failed during implementation.

Nambisan and Wang (1999) suggested that the average organization is still finding it difficult to address some of the basic issues related to Web technology adoption. The adoption of new technologies is dictated by factors such as perceived costs and benefits, complexity, compatibility with existing systems, and ease of use. An adopting unit may face knowledge barriers that may delay or otherwise affect the adoption process adversely. Technology-related knowledge barrier relates to the lack of knowledge regarding the appropriate hardware and software infrastructure, technology features, security, and standards. Project-related knowledge barrier includes the lack of knowledge regarding resource requirements for Web-based application development, development process and duration, project leadership, and functional participation. Application-related knowledge barrier relates to the lack of knowledge regarding the specific business objectives that will be served by the Web-based application, the value of the various technology features for the adopting unit, the key business assumptions required to be made for deploying the technology, the potential for integrating the application with existing IT applications, and the effect of the Web application on current organizational structure and systems.

USDA - NRCS (1995) examined all aspects of data and information collection, analysis, and dissemination within the NRCS; and reported that the agency's broaden mission and information needs have not been recognized by many NRCS personnel. Two types of issues were identified at NRCS. One type revolved around management and agency cultures—questions about roles of different levels of the agency, priority-setting, decision-making processes, and the influence of personal values and views. The second type revolved around technical and technology topics—questions of hardware and software compatibility, use of commercial software, frequency of technology upgrades, and staff responsibilities in applying modern technologies.

It is feasible to deliver up-to-date online maps and attributes efficiently and effectively using dynamic update techniques.

Limitations

Internet-related technology has evolved very fast; it is not easy to keep up with seemingly constant changes and developments in software, hardware, and standards. This study focused only on current, proven technologies that are relevant to the development and implementation of dynamic update techniques.

Definition of Terms

cartographic data	Data relating to the making of maps or charts.
digital hydrologic unit (HU) data	Digitized data defining the locations of river basins, watersheds, and sub-watersheds.
digital orthophotography quadrangles (DOQ) data	Digitized data showing distortion-free aerial photographic images that are formatted as standard 7.5-minute quadrangles (15-minute in Alaska) or as quarter-quadrangles at a scale of 1:12,000.
document data	Refers to data or files created with a word processor, such as MS-WORD.
geographic information system (GIS)	A configuration of computer hardware and software that stores, displays, and analyzes geographic data.
geospatial data	Spatial data representing the shape, location, or appearance of geographic objects. It can be vector, raster, or image format.
map unit interpretation record (MUIR) data	A collection of soil and soil-

	related properties, interpretations, and performance data for a soil survey area and its map units, map unit components, and component layers.
metadata	A definition or description of data.
natural resources inventory (NRI) data	Data on the status, condition, and trends of the nation's soil, water, and related natural resources.
soil survey geographic (SSURGO) data	Data used in detailed soil survey maps. SSURGO data are linked to Soil Interpretations Record attributes to give the proportionate extent of the component soils and their properties for each map unit.
soil survey laboratory characterization (SSL) data	Data representing the central concept of a soil series or map unit sampled to bracket a range of soil properties with a series or a landscape.
state soil survey geographic (STATSGO) data	Data used in state soil maps and made by generalizing the detailed soil survey or SSURGO geographic database.
remotely sensed image data	Data acquired from a distance by satellite imagery or aerial photography.

status map — A map that symbolizes features according to the state or condition of a particular attribute at a particular time. For example, a U.S. map displaying the status of archived SSURGO data in different colors.

tabular data — Refers to RDBMS textual (American Standard Code for Information Interchange, a.k.a. ASCII) data containing human-readable information.

topographic data — Data describing the three-dimensional shape of a land surface, including its relief and the position of features.

Summary

There is a need to accommodate perpetual data updates in the database and the delivery of real-time online maps and attributes information. The goal of this dissertation was to analyze and evaluate dynamic update techniques and technology. The result of this analysis culminated in the design and prototype of a system that allows the real-time delivery of up-to-date maps and attributes information.

Chapter 2

Review of the Literature

Historical Overview

In mid-1997, NRCS began to envision field conservationists being connected to the world electronically via the WWW and accessing data and information in real time. The vision called for a new direction in information technology that will allow better, more accurate, and timely information, using as much commercial off-the-shell (COTS) software as possible (USDA - SC, 1997). The NRCS Soil Survey Schedule (SSS) Design Team, for example, met in January and February of 1998 to analyze the critical business requirements of the Soil Survey Program and to design an implementation strategy to meet these needs. The design team received approval in May 1998 for their SSS Rapid Application Development (RAD) Implementation plan to develop a Web application for users in the field. This application provided, among other things, the capability to report mapping progress, to track the progress of various survey products, and to manage and track compilation and digitizing for SSURGO (USDA - NRCS, 1998).

The USDA - SC (1998) described similar requirements in its strategy to integrate GIS technology and business reengineering processes. Besides identifying several critical geospatial data requirements to support the service centers' mission, it singled out the importance of optimizing data storage and delivery and minimizing costs and redundancy while maximizing data integrity, delivery efficiency, and system performance. The Geospatial Data Acquisition, Integration, and Delivery Business Process Reengineering (BPR) Project was created to address those issues. The BPR project specified the goals of access and delivery to include, among other things, support for more efficient and timely program delivery with Internet Web, data servers, and FTP sites.

The UCGIS (1996) also cited a requirement in updating attribute data. They described the need to develop methods that are capable of extracting and updating essential metadata, which is a shared system for describing data and is a key component of any interoperable environment. The Internet was described as a breakthrough that offers the opportunity to integrate data from widely different sources and makes it easier to process geographic information. For this case and the two cases previously described—whether the requirement was for timely and accurate accountability information, optimal delivery of geospatial data, or the capability for online update of metadata—the desirable solution was a Web-based updating system.

Adam et al. (1996) gave an insight into why dynamic update technique is important to online data. They described digital libraries (DL) as being dynamic, in that materials can be added and updated by many authors; and for electronic commerce (EC), price lists, advertisements, and the like will change over time. An EC/DL system requires methods to facilitate the update of existing content, and to capture continuous digitized media in real time.

Roehle (1997) agreed with Adam et al. (1996) that changes instigated by a user on one system must be communicated to other users on the network. He also added another reason why dynamic update is so essential. In a distributed virtual environment (DVE), "this updating allows users to have a consistent view of the world: when a user opens a particular door, all of the others with an unobstructed view should see it open" (p. 32). It will be demonstrated below in the survey of the types of dynamic updating techniques currently in use, that dynamic update techniques are increasingly necessary for online data. In summary, substantiating reasons for using dynamic update techniques for online data include: the need for technology that will allow better, more accurate, and timely information; the importance of support for more efficient and timely program delivery; and the need for methods that are capable of extracting and updating essential metadata.

Theory and Research

Types of Dynamic Updating Techniques Currently in Use

Numerous accounts of dynamic updating technology and techniques can be found in printed and online literature. They can all be grouped into three types of Web-based dynamic update techniques in use today that are designed for online data:

- Dynamic HTML using Document Object Model (DOM) technology, Common Gateway Interface (CGI), Java, JavaScript, XML (eXtensible Markup Language), or ActiveX.
- Internet Object Pushing (a.k.a. WWW Push)
- Application Programming Interface (API)

The EXPRESS Web Server (Sauder & Lubell, 1999) presents an example of the first type of dynamic update technique using CGI, an established server-side technology. The Server allows users to generate Web pages dynamically using a CGI script. User-directed input from a Web page leads to the dynamic creation of another Web page where files on the Server are added and deleted. Although the EXPRESS Web Server uses a technique that retrieves data from a repository, the data are static in nature.

The Active View system (Abiteboul, 1999) developed at INRIA (France) presents another example of the first type of dynamic update technique using XML and Java. In Active View, the database server is an XML repository and the view server is a Java application. The XML repository exports XML data and the view server restructures the data to construct the view. The XML view document, which is in dynamic HTML with embedded Java applets, is handled by a standard Web browser and interacts with the view server to obtain data. Active View enables the declaration of an active rule to specify that, when a notification of change occurs, the view should be updated. The focus in the Active View system is on the control of updates and on the declarative specification of view.

The Network User interfaces (NUIs) (Neerincx et al., 1999) present an example of the second type of dynamic update technique. Currently being developed by Telematics Institute (Netherlands), the NUIs offer a browser-like interface for navigating through local and remote file systems. They can display and automatically update dynamic Web content using "push" and "pull" Webcasting technologies.

Padula and Rinaldi (1999) describe a work that presents an example of the third type of dynamic update technique. The work addresses the communication needs of seismologists, allowing them to browse all available information on seismic events, including parametric earthquake records, original texts, and intensity maps. Because of the social relevance, seismological operators must rely on the quality of and easy access to the stored data in real time. The environment is composed of dynamic Web pages that access the constantly updated Database of Macroseismic Observations (DOM4.1) and National Group for Protection of Earthquakes (NT4.1) databases. SQL (Structured

Query Language) and PHP/FI (Personal Home Page/Form Interpreter) languages make it possible to rapidly access data that are stored and managed with the CDS/ISIS system developed by UNESCO.

This short survey shows that dynamic update techniques can be found on systems that operate on either unchanging or constantly changing data. They are increasingly integrated into online data systems, and are transparent to users.

What is a Dynamic Update Technique?

In their paper presenting the BUS (Bandwidth-sensitive Update Scheduling) method for WWW push proxies that actively sets different update schedules for various WWW push channels, Huang and Yu (1998) describe dynamic update as "a mechanism to monitor client interest and to conduct a dynamic proxy update for given channels if the client requests for these channels increase suddenly; this mechanism can better capture dynamic changes (such as financial activities) that are of interest to clients" (p. 303). This description is based on WWW push where information is continually sent from the servers to a large number of clients who set their preferred update schedules beforehand.

Plasil, Balek, and Janecek (1998) base their description of dynamic update on the SOFA (SOFtware Appliances) component model and its extension, DCUP (Dynamic Component UPdating), with two important new emphases. For them, dynamic update designed for SOFA/DCUP has to take place "with minimal effort/interaction at the end-user side," and "at runtime if necessary, e.g., in real-time applications" (p. 43).

The new emphases accentuate technologies that (1) minimize user intervention in receiving updates, and (2) allow the updates to occur in real time. The first emphasis is important in every dynamic update technique. In the BUS method, user interaction is minimized to a certain degree, being limited to the individual selection of pre-specified channels from which to receive updates. The second emphasis is on promoting the delivery of real-time, up-to-date information. This is especially important because this aspect of dynamic update is actually implemented in systems such as those described by Caputi (1998), and Padula and Rinaldi (1999).

Developing a Dynamic Update Technique

Huang and Yu (1998) describe the development procedure of a dynamic update technique for the BUS method as the following:

- Set a dynamic update checkpoint interval.
- At every dynamic update checkpoint, retrieve the gateway traffic information from the network devices in order to measure the percentage of bandwidth used for all traffic in the last time interval.
- Estimate the percentage of the bandwidth available for dynamic updates for the next time interval.
- Select the channels with the highest non-zero obsolescence sums.
- Perform a dynamic update for each of the selected channels.

This procedure focuses on a WWW push strategy, which optimizes the overall currency of the push objects while constraining the overall push traffic to the available bandwidth. It resembles the general procedure for systems that use the WWW push strategy in their dynamic update technique.

In contrast to the above procedure, Bromberek and Medina (1998) report on the unique dynamic update capability of INSIGHT, an electronic publishing product made by

Enigma that integrates document management solution and enables the use of Word and XML for data reuse. INSIGHT's dynamic update technique implements two essential features. The Update Wizard pushes simultaneous updates out to end users while supporting the addition, deletion, and modification of published CD-ROM data. The Scheduler module creates task lists that dynamically update to the production process at scheduled intervals. For example, the Scheduler can be instructed to check for updated files once every hour; if updated files are found, it will then automatically run the build wizard to incorporate the changes.

Huang and Yu (1998) emphasize the push method as the choice for their technique. It is important that the developer of a dynamic update technique decide early on the method. The method may be dynamic HTML, Internet Object Pushing, or API. Based on the chosen method, the developer may decide on the language and technology during integration, which may be DOM, CGI, Java, JavaScript, XML, ActiveX, or a combination of several of those. Attention to the method is important in the development of a dynamic update technique and is vital to the success of the system implementation.

Huang and Yu (1998) also pay special attention to bandwidth—a critical issue for online updating systems that must transmit great amounts data over the network. The emphasis on bandwidth is important because the goal of their method is to minimize the push traffic overflow and to better capture the dynamic changes.

The techniques are designed with different emphases in each case. Huang and Yu (1998) develop a push method that is based on the historical push access pattern and the knowledge of each channel to compute an update schedule for each channel while constraining the overall push traffic to the available bandwidth. The method consists of three main tasks. One task handles client requests. Another task conducts periodic updates for each subscribed channel. The third task dynamically updates certain channels if the level of client interest for these channels shows a dramatic increase. This method effectively optimizes update traffic based on the available bandwidth.

With INSIGHT, Bromberek and Medina (1998) describe the development of a method that features various wizards to assist in the management of content and updates, and in the scheduling of dynamic updates. INSIGHT architecture includes NetSight—a combination of HTML forms, JavaScript, and CGI—that searches the database, returns query results, dynamically converts requested documents to HTML, and passes them back to the user. CGI, however, is considered first-generation technology that typically runs slowly and has difficulty handling large numbers of concurrent requests.

More recently, Brown (1999) describes the development procedure for a prototype virtual reality user interface (VRUI) that is intended to allow dynamic interaction with geographic data in the context of a virtual environment. The Web pages use frames and display an initial contents frame listing available areas that, once selected, load a 3D virtual scene embedded into the page by the plug-in and locally installed VRML browser. The 3D scene is a JAVA applet with a menu of buttons used in conjunction with the 3D navigation tools of the native VRML browser (e.g., zoom, pan) to control the scene contents. A button in the VRUI provides a link to extract and display metadata information to the user. A VRML enabled time-sensor runs as a background process to incrementally load new data at set time intervals. An important feature of the technique in this procedure is that new data sets and temporal updates can be periodically added.

The coupling of VRML and JAVA in this technique provides an innovative new means to query and browse spatial data. There is one challenge in this technique, and that is the issue of cross-platform compatibility. VRML browsers can interpret the current specification differently from one another; and Web browsers differ in their implementation of some JAVA classes. Tighter standards need to be developed before a more widespread implementation of this technique can be seen.

Online Delivery of Maps and Attributes Data

Because dynamic update techniques are becoming increasingly integrated into online data systems, interests have risen about implementing dynamic updating capability in online systems that deliver maps and attributes data. Newton, Taylor, Trinidad, Ackland, and Abel (1999), on the subject of Information Integration, state that "the Internet driven demand for immediate, online access to digital data is challenging our abilities to organize, structure, and present information in ways that are useful to both casual and experienced end-users" (p. 91). They mention the Sydney Water's Hydra5 project that manages a wide variety of digital data, including textual material, graphical material (aerial or site photographs), land use and infrastructure maps for proposed or actual scenarios, input and output files for simulation models, real-time telemetry data, precipitation records, and biodiversity survey results. Central to the Hydra5 data model is architecture supporting rich hyperlinks between spatial entities, maps, descriptive text, metadata, attribute data, and arbitrary digital files. It supports spatial browsing to locate information, which involves the linking of entities, such as the current environment status and predicted rain events; and effectively integrates spatial and non-spatial information.

On the subject of Internet GIS, Harder (1998) states that geographic information that changes over time, such as weather patterns, can be shown with frequently updated maps. An example was given of the Weather Channel's Web site where a satellite weather map is updated hourly with a fresh image beamed down from low earth orbit. The maps are served as embedded GIF or JPEG images. A script running in the background replaces the image whenever a new one becomes available.

In similar views, Balasubramanian and Bashian (1998) state that Web information system (WIS) designers "must resolve issues of authoring, organizing, managing, and delivering large amounts of unstructured and timely information via the Web" (p. 107). Bryron, Kenwright, Cox, Ellsworth, and Haimes (1999) discuss the payoff of a responsive, intuitive virtual reality interface that enables the use of visualization tools to query and explore spatial data quickly. Plaisant and Jain (1994) discuss the use of dynamic queries on a thematic map application that displays spatio-temporal statistics. They conclude, "these methods hold much promise for improving access to large data sets" (p. 439).

Summary

The literature review shows that there is a need for Web-based updating systems and technology that allow efficient and timely delivery of information, such as accountability, geospatial, and metadata information. The review also shows that dynamic update techniques are increasingly integrated in online data systems; and that interests exist in implementing dynamic update capability in online systems that deliver maps and attributes data.

Contribution

An ample amount of research material exists on the theory and practice of developing dynamic update techniques. Despite that fact, no research literature was available that specifically deals with dynamic update techniques that provide for real-time delivery of up-to-date maps while allowing online update of attributes information. This dissertation was the first attempt at providing research material in this important area.

Databases searched: ACM Digital Lib, IEEE Web sites, Nova Southeastern University Distance Library system, Netscape, Yahoo.

Keywords used: Dynamic update, dynamic updating, data storage, data delivery, data access, geospatial data, real time, online access, access method, up-to-date information, update technique, online update algorithm.

Chapter 3

Methodology

Research Methods

Earlier in the dissertation, the types of dynamic updating techniques currently in use were discussed. It is time to integrate dynamic update techniques in a delivery system for online maps and attributes data. To achieve this integration goal, an action research method with formative evaluation (Mauch and Birch, 1993) to the investigation was applied. This chapter provides details of this approach.

Specific Procedures

To accomplish the stated goal, the first step was to study data collection and storage techniques. NCGC, being a data clearinghouse, handles many kinds of data in great quantities. As mentioned above, these data come from many sources and in many forms. The study included how this conglomerate of data is being collected and stored by various departments within NRCS. The key factors in evaluating storage needs included annual storage, retrieval, and update requirements; anticipated growth; required access times; and file and image size (Grare, 1995). This dissertation examined tools that employ more efficient storage (Kirstein, 1996), and alternative retrieval and access techniques, such as signature files (Croft & Savino, 1998; Faloutsos, 1985), transient hypergraph-based model (Watters & Shepherd, 1990), and database optimizers (Mackert & Lohman, 1989). New approaches to solving the problem of data management system, such as multidatabase or federated systems (Litwin, Mark, & Roussopoulos, 1990), were updating the entire data set. Users would not accept unnecessary waits when only parts of the data set are being updated, or when some of those data on the client side can be refreshed or processed. If the delay is too great, the system becomes difficult or even impossible to use (Roehle, 1997). This dissertation proposed to examine how to improve the utilization of main memory systems and how to apply these approaches under update traffic (Kamel & King, 1992). The use of Java applets, ActiveX controls, ESRI shape file, and emerging technologies for client-side processing and updates were analyzed. Data security issues were also investigated.

The fourth step was to assess the feasibility of hardware, software, user interface, databases, Internet initiatives, tools, and technologies germane to collecting, storing, and updating techniques. They must meet the cost and usage requirements of the present and the future for optimum infrastructures that can be applied in typical situations, such as NRCS.

With the knowledge gained in those studies, the succeeding steps were to design a detailed plan for these techniques, to apply the newly conceived plan in selected test situations, to evaluate the results, and to propose new challenges. SSURGO is one type of data that NCGC has both quantity and quality controls of; hence, SSURGO data and their archived status were the subject of the information system prototype to be developed in this study. Preliminary analysis had determined the needs for tabular and spatial representations of SSURGO status information. The building blocks of the pyramid model (Whitten et al., 1994) for conceptual foundations of the system are described in the following paragraphs.

explored. Following are some suggestions in literature applicable to NCGC regarding data collection, storage, and access. Wilensky (1996) suggested that servers be repositories implemented as databases supporting user-defined functions and user-defined access methods. New user interface paradigms and improved protocols will be needed for client-program interaction with repositories (collections). Faundeen and Zanter (1997) provided guidance on the management of metadata information, which can be applied to the dynamic data collection that are updated on a daily basis and are made accessible to the public. Orland, Wu, and Chavan (1997) ascertained that Web delivery would be the best method to serve significantly large repository of up-to-date information.

The second step was to study online data delivery techniques. This dissertation proposed to examine telecommunication areas, such as bandwidth, connectivity, and interoperability. Techniques for improving online performance, such as runtime partitioning and scheduling where a user request is transferred to client-side machines to reduce network traffic (Andresen, Yang, Ibarra, & Egecioglu, 1998), were studied. Dynamic Web database tools, such as Microsoft SQL Server, NetObjects Fusion, ESRI Internet Map Server, and Apple WebObjects were examined. Advances and the pace of change in these areas are rapid and numerous. The focus of this research was on fast, effective, reliable, and practical techniques for data delivery that adapt to technological changes, and optimize resource utilization.

The third step was to study techniques for real-time data update. In general, online data updates require round trip messaging from the client to the server. This type of messaging may be acceptable when transmitting relatively small chunks of data or when

Subject matter experts (SME) at NCGC, including those who are responsible for daily data collection and dissemination, were consulted face-to-face. Other SMEs at NRCS state offices and digitizing units were also consulted by electronic means (i.e., telephone, fax, and email). Currently, the NCGC - Automated Mapping Section (AMS) uploads SSURGO data when state offices and digitizing units send email notifying them of the updates. The NCGC - SSURGO Support Section validates the data prior to putting them in archive. The NCGC - AMS also compiles status information of the archived SSURGO data. From the SSURGO repository, NCGC makes available on the Internet data for public downloading via FTP, a tabular status list, and a spatial status map in JPEG (Joint Photographic Expert Group) format. The technology goal was to automate these manual processes, making them more effective and efficient. The system used existing, emerging technology, and optimal integration techniques to automate the processes of uploading data from state offices and digitizing units, validating data, transferring them to the FTP site, and updating the tabular status list and spatial status map.

The OMT (Object Modeling Technique) methodology (Rumbaugh et al., 1991) was used during analysis and design. A statement of needs, including performance specifications and interaction protocols, was compiled for requirements. An object model describing object classes and their relationships was constructed based on the results of the requirements analysis. A data dictionary for all modeling entities was prepared. Association between classes, object attributes, and operations was identified and refined. A dynamic model was constructed as necessary to show the time-dependent behavior of the system and the objects in it.

A relational database, which is based on the refined object model, was implemented following the design methods outlined by Elmasri and Navathe (1994). This database is served by an integrated ORACLE RDBMS (Raphaely, 1997), which is the database server of choice by the agency. Programming in HTML (Castro, 1996), JavaScript (Danesh, 1997; Kent & Kent, 1997), Java (Cornell & Horstmann, 1996; Hoff, Shaio, & Starbuck, 1996), C++ (Horstmann, 1997), Visual BASIC (McKelvy, Spotts, & Siler, 1997), CGI Perl (Deep & Holfelder, 1996), and Unix Shell (Swartz, 1990) were carried out appropriately.

OTS (Off-The-Shell) software tools were evaluated to best meet the design and integration requirements of the client-server architecture and existing equipment at NCGC. Training or any other resources deemed necessary to complete the project were identified and requested.

Chronologically Ordered Listing and Summary of the Procedures

Step 1: Data Collection and Storage

With emphasis on dynamic data:

a. Gathered information from NCGC data stewards and team leaders on how those data are being collected, uploaded, validated, and stored; and on the status of their updates.
b. Discussed with concerned personnel the possibility of simplifying and automating the current routines (i.e., data collecting, uploading, validating, archiving, storing, and updating).

c. Proposed a persistent, robust procedure in addition with programs and software to support the above data functions more advantageously.

d. Devised a method to convert the existing in-house data as a part of the overall integration plan.

Step 2: Data Delivery and Access

a. Investigated technological advances in online delivery and access of spatial data.

b. Evaluated those technologies and determine whether any of them was an appropriate solution for NCGC and its infrastructure.

Step 3: Update

a. Studied current techniques used for real-time data update.

b. Determined and developed an update technique that operates seamlessly on dynamic data.

Step 4: Assessment

Determined the feasibility of implementing any of the technological solutions.

Step 5: Prototype

a. Engineered an object-oriented design of the decided technological solution.

b. Procured necessary software, hardware, and associated training.

c. Performed required programming.

d. Tested the prototype.

Formats for Presenting Results

SMEs at NRCS state offices and digitizing units, and NCGC SMEs including section chiefs, team leaders, and employees involved in daily data collecting and disseminating activities were the primary source of information regarding data collection and storage. These people were the most knowledgeable about the problem and were the most interested in this research. Amongst them, managers were concerned about the accountability status of the data, and others were interested in reducing the burden of work that is required to keep up with all the changes. For each type of dynamic data, questions such as the following were asked, not necessarily in this order:

- How do the data get here?
- How are they being collected in the beginning?
- What tool, if any, is being used for the data collection?
- What does NCGC do with them?
- What is the frequency of data updates?
- Who is validating them?
- Where do the data reside?
- Who is requesting them?
- How are they being distributed?
- What are management concerns?
- What are data keeper's concerns?

Primary channels for communication with the SMEs and others included telephone, fax, email, and meetings. Brainstorming sessions were held as necessary to work out details and to overcome technical, procedural, and organizational obstacles. Simple white boards and the more high-tech SoftBoards were available in those meetings to help conceptualize ideas. The hardware-software devices (e.g., SoftBoards) made it easy to sketch plans, diagrams, or flowcharts; and to print hardcopies for reviews.

The Internet was a tremendous aid in the investigation of existing and emerging technologies. It provided the fastest way to a wealth of information on technology news and products. The majority of software available for evaluation could be downloaded or requested while being on the Web. In addition, most of the evaluation software was available for two popular platforms: UNIX or Microsoft operating systems (OS). NCGC houses powerful SUN workstations running Solaris, which is a form of UNIX; and Windows NT personal computers (PC) and laptops. Most contemporary language compilers, interpreters, and IDEs (Integrated Development Environments) were available on, or easily obtainable for, both of those platforms. This fact was important because a significant amount of programming was anticipated.

Microsoft (MS) Office 97 products (e.g., MS-Word, and MS-Excel, etc.) and a Visible Systems Corporation's Computer-Aided Software Engineering (CASE) tool, Visible Analyst, assisted in the OO analysis and design of the prototype. Products from these software tools included requirements statements and various object models, copies of which were distributed for reviews and during meetings.

One or more network and application performance tools, such as VitalSoft (Lucent Technologies - Lucent NetworkCare [LT - LNC], 2000), were used to monitor and measure the performance of the prototype. These tools helped assess how well the proposed solution met with continuing interests and concerns, in addition to how well it responded to the needs and requirements, of NCGC.

Outcomes

The results of the study supported the original hypothesis that it is feasible to deliver up-to-date online maps and attributes efficiently and effectively using dynamic update techniques.

Resource Requirements

NCGC had agreed to sponsor the proposed dissertation work because it had significant relevance to its mission and is of great interest and importance to the departmental and sectional managers. Because NRCS Information Technology Institute (ITI) technology directions affect NCGC, the directors at ITI and NCGC agreed that ITI provide technical guidance for the work. Additionally, this dissertation also derived some benefit from a collaborative project that ITI has with the Department of Urban Studies and Planning at MIT (Massachusetts Institute of Technology). The project with MIT was for technologies that facilitate the serving of USDA geospatial data.

Sophisticated hardware and software support was available for research from NGC's Information Systems Support (ISS) group. Some of the emerging technology software support came from ITI. High-speed (T1 capability) Internet access was also readily available.

The NRCS facility in Fort Worth has a library with much GIS-related literature and many books and journals on information and technology. Local public and university libraries were conveniently accessible. Besides the Nova Southeastern University Distance Library system, the author also had members-only access to the Association for

Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE) digital libraries.

Reliability and Validity

Validating the reliability and effectiveness of the dynamic update technique involved performance and usability studies. One of the tools from the aforementioned VitalSoft (LT - LNC, 2000) was used in the performance study to provide detailed network and application performance metrics. The tool provided charts displaying the performance of application transactions on the network; and reports showing network transaction times, throughput, latency, and other network related measurements. A study of this metrics demonstrated whether the response times were in acceptable range.

Reviewers participated in the usability study of the implementation by answering questions such as:

- How is the graphical presentation or design? (For example: placement of menus, buttons, and objects; cluttering; etc.)
- How well does it retrieve the information you need?
- Do all the buttons and links work as expected?
- Are you comfortable using it?
- Is update information correct?
- Is the wait appropriate and acceptable for graphics intensive uploads?

The answer to these types of question verified whether the implementation meets the up-to-date information requirements.

Summary

The procedure consisted of five major steps encompassing a number of small steps, and culminated in the development of a prototype. The steps included gathering data collection and storage information, investigating technological advances in data delivery and access, studying dynamic update techniques, assessing the feasibility of an implementation solution, and developing a prototype. The results revealed that the dynamic update technique as implemented in the prototype met the need for timely delivery of accountability, geospatial, and metadata information within NCGC infrastructure.

Chapter 4

Results

FINDINGS

Step 1: Data Collection and Storage

A number of preliminary meetings with NCGC data stewards, team leaders, and SMEs were conducted to examine the current business processes that were being used in the collection, uploading, validation, and storage of geospatial data. A sample of the questionnaire used to initiate discussions is listed in Appendix B. A majority of the types of data that are being served at NCGC was discussed, including SSURGO, STATSGO, NRI, DOQ, MUIR, and SSL. SSURGO was singled out as the type of data that undergoes most frequent changes and updates.

Several spatial and thematic issues were discussed with concerned personnel regarding SSURGO (see Appendix A - Correspondence - "SDE Meeting Notes and Action Register"). Spatial issues included frequency of digitized data uploading, archive location and storage requirements, files content and format, and method of delivery. Thematic issues included personnel involvement in updating the status of soil survey digitizing map. The consensus was for software implementation to focus on (1) simplifying and automating the procedures to upload, validate, archive, update, and deliver digitized SSURGO data sets; and (2) automating the SSURGO digitizing status map updates.

A Digitizing Unit (DU), of which there are seven, is an office that is responsible for digitally capturing soil delineation from a soil survey. They certify the data, along with the MLRA Office, and send to NCGC for archiving. Procedures were established to simplify the uploading, to coordinate the validation, and to automate the archive, update, and delivery of SSURGO data coming from DUs. These procedures relied heavily on UNIX shell scripts that query the data and other files for FTP posting and distribution. For the SSURGO digitizing status map, a dynamic Web technique was used to accomplish the updates.

Step 2: Data Delivery and Access

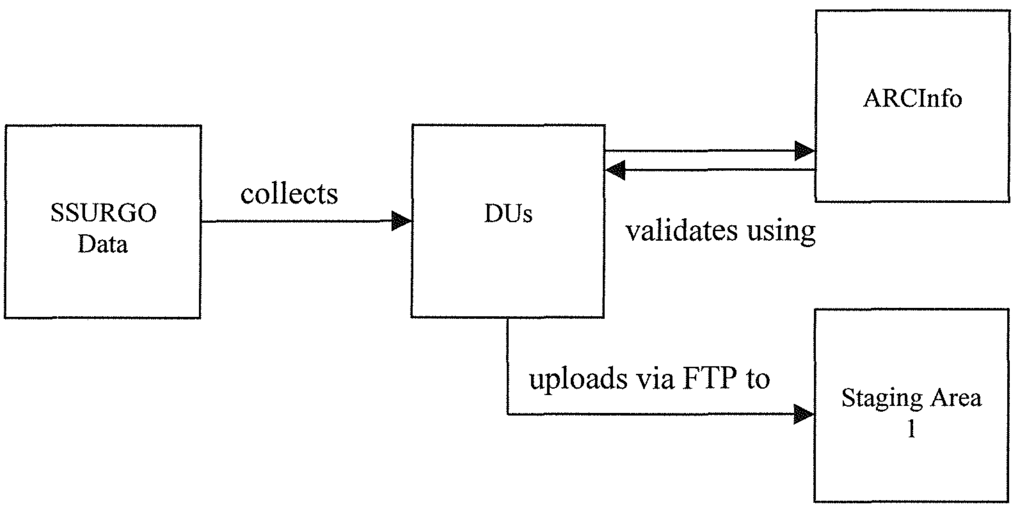


Figure 1. Data delivery and access.

The DUs continue to collect SSURGO data from the field. They use ArcInfo, a GIS product from ESRI, to validate the data. Once the data are validated, the DUs upload them via FTP to a private location (a.k.a., Staging Area 1) at NCGC, which is an improvement from the past when SSURGO data arrived after several days at NCGC via 8mm tapes. Appendix C lists the procedures for validating certified SSURGO Data via FTP.

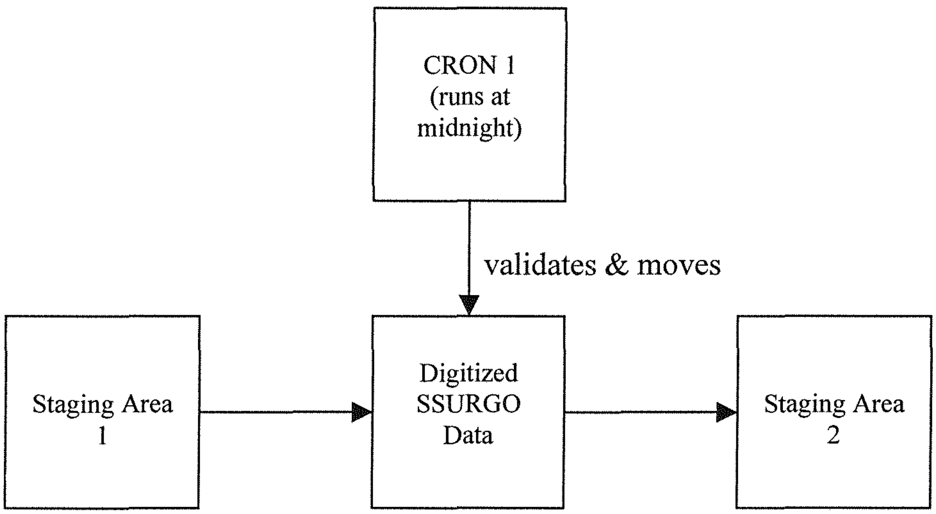


Figure 2. Validation and move.

A primary Unix shell script was written and set up as a CRON job to run at midnight everyday. It picks up all newly uploaded SSURGO data sets from the DUs' private area, Staging Area 1, and validates the data before moving them to Staging Area 2. A SSURGO data set consists of the following files:

- cov.zip - SSURGO ArcInfo coverage zipped file
- dlg27.zip - DLG in NAD27 datum format zipped file
- dlg83.zip - DLG in NAD83 datum format zipped file
- tab.zip - Tabular attribute zipped file
- out.zip - ArcInfo output zipped file

- stssaid.met - Metadata file
- readme

The validation includes checking for incomplete or corrupted data sets, file name mismatches, etc. The script sends email to DUs and other concerned personnel to inform them of the failure or success status of the data validation and transfer (see Appendix A - Correspondence - "SSURGO data set successfully received"). A successful validation results in the data set being moved into Staging Area 2. A failed validation results in the data set being left on Staging Area 1 for ensuant corrections by the DUs.

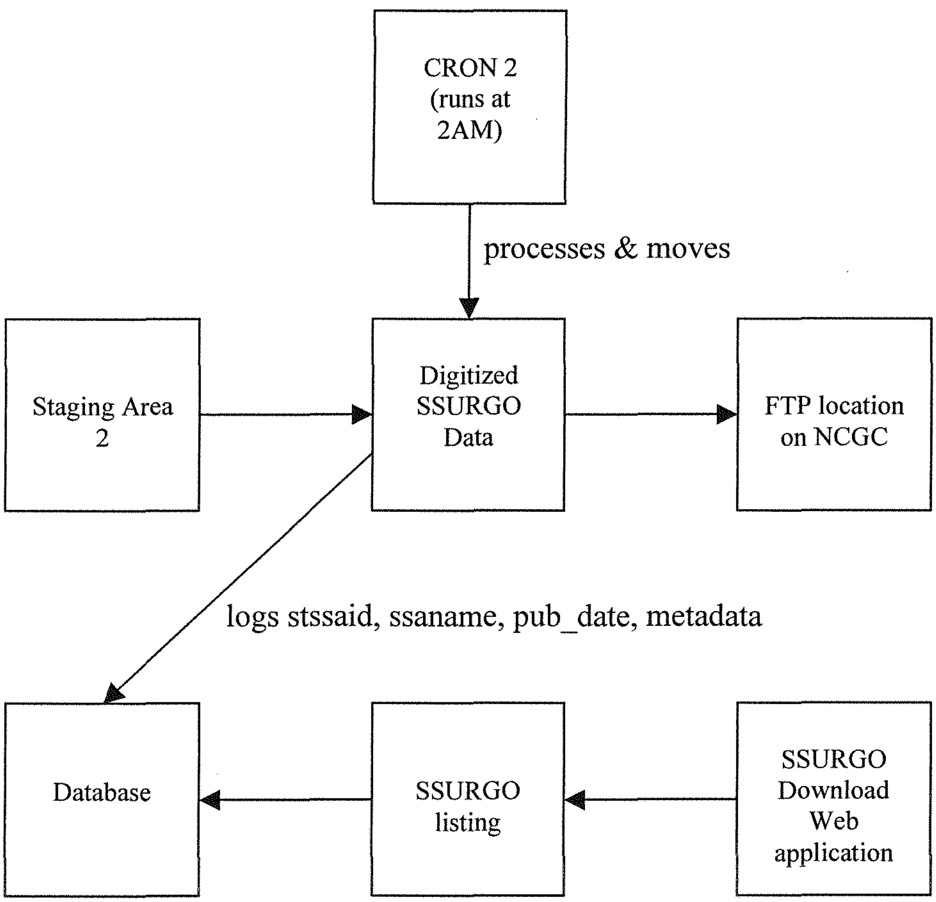


Figure 3. Process and move.

Another primary Unix shell script was written and set up as a CRON job to run at 2AM everyday, thus allowing plenty of time for the other CRON job that is started at midnight to finish. The script parses through Staging Area 2 to look for all newly added and validated data sets that have been put there by the previous CRON job. Here, a database record is logged for each data set with unique identification information including soil survey id, name, publication date, and metadata file name. Other processing includes unzipping files, checking and special handling for revised data sets, etc. The script then moves data sets to the FTP site on NCGC and makes them available for public downloading. Before the script exits, it sends an email out to concerned personnel informing them of the move (see Appendix A - Correspondence - "SSURGO Data ... moved into FTP location"). It also calls an ESQL/C program to dynamically build a listing of available SSURGO data sets by querying the database for logged soil survey entries. This listing is used in a Web database application named SSURGO Download (http://www.ftw.nrcs.usda.gov/ssurgo_ftp3.html) that is heavily accessed by the public. Following are some screen captures of the main pages of this application:

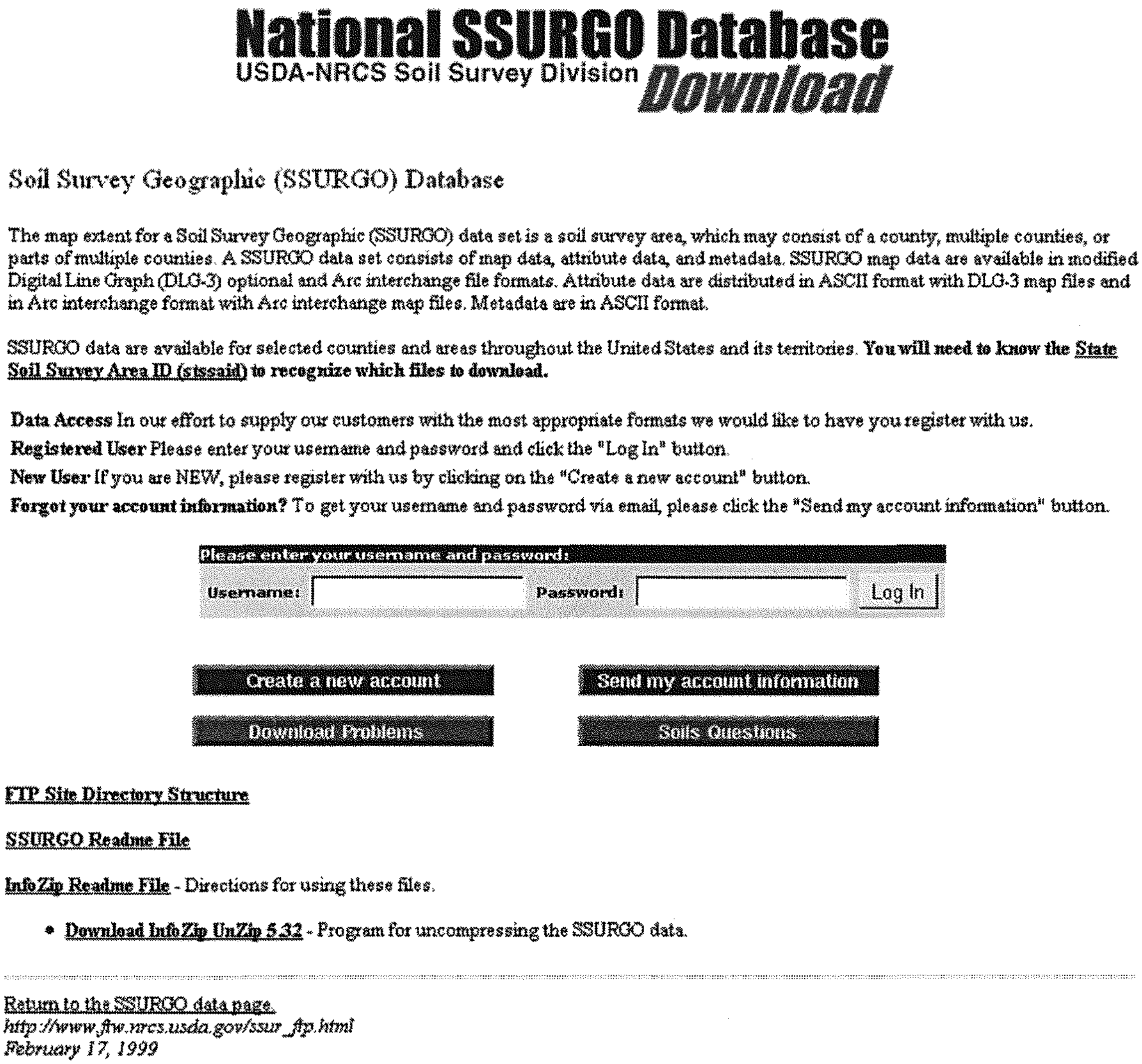
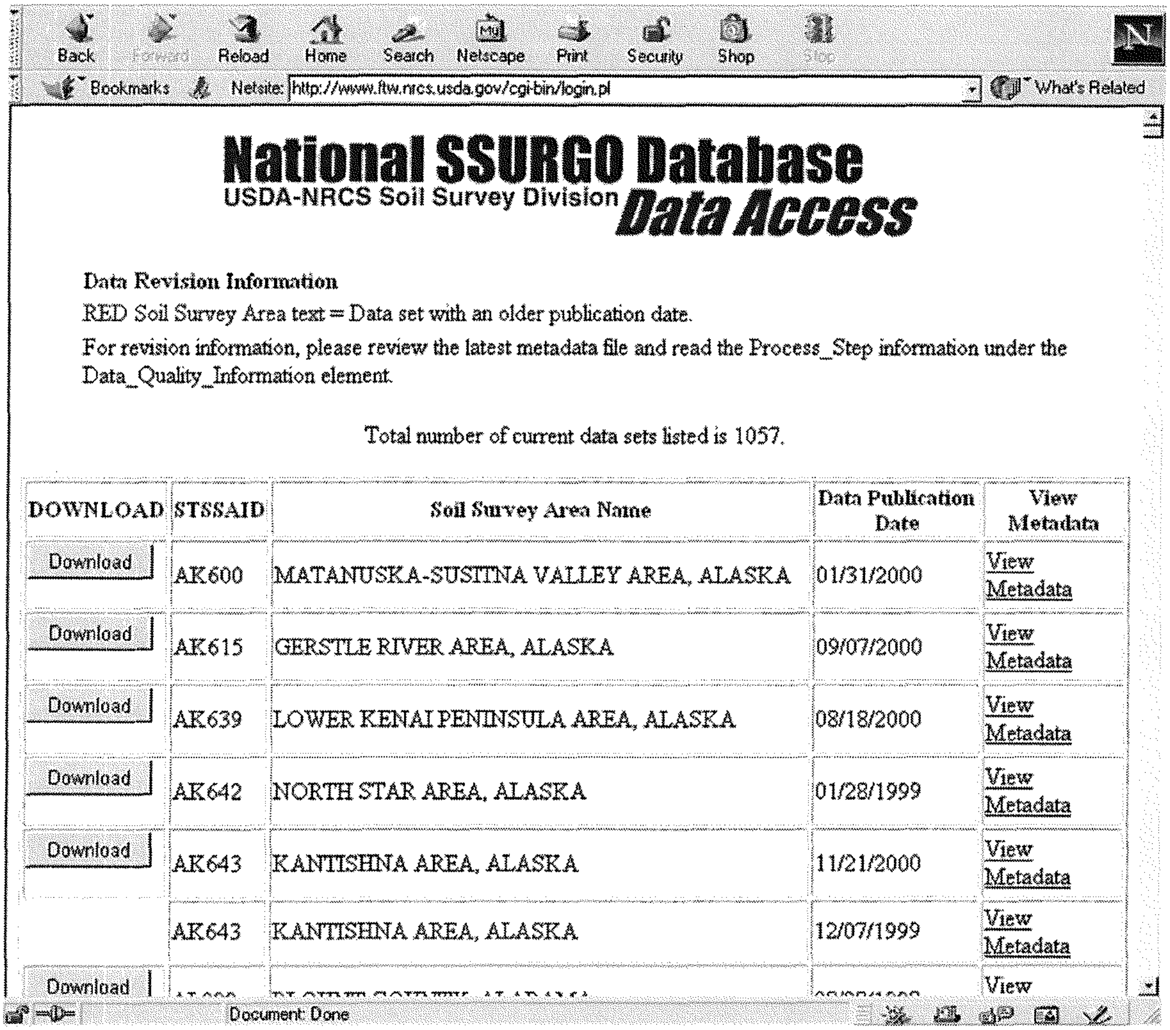


Figure 4. SSURGO Download application.

Figure 4 shows the first page of the SSURGO Download application. Users must be registered in the database and have their account information set up before they can log in to use the system. Although the data are not regulated or restricted from any user, this information is necessary for the generation of accountability reports to management.



National SSURGO Database
USDA-NRCS Soil Survey Division
Data Access

Data Revision Information

RED Soil Survey Area text = Data set with an older publication date.

For revision information, please review the latest metadata file and read the Process_Step information under the Data_Quality_Information element.

Total number of current data sets listed is 1057.

DOWNLOAD	STSSAID	Soil Survey Area Name	Data Publication Date	View Metadata
Download	AK600	MATANUSKA-SUSITNA VALLEY AREA, ALASKA	01/31/2000	View Metadata
Download	AK615	GERSTLE RIVER AREA, ALASKA	09/07/2000	View Metadata
Download	AK639	LOWER KENAI PENINSULA AREA, ALASKA	08/18/2000	View Metadata
Download	AK642	NORTH STAR AREA, ALASKA	01/28/1999	View Metadata
Download	AK643	KANTISHNA AREA, ALASKA	11/21/2000	View Metadata
	AK643	KANTISHNA AREA, ALASKA	12/07/1999	View Metadata
Download	[illegible]	[illegible]	[illegible]	View

Figure 5. SSURGO listing page.

Figure 5 shows the SSURGO listing page displayed after a user successfully logs in from Figure 4. This listing is re-generated automatically every night with updated information from the database.

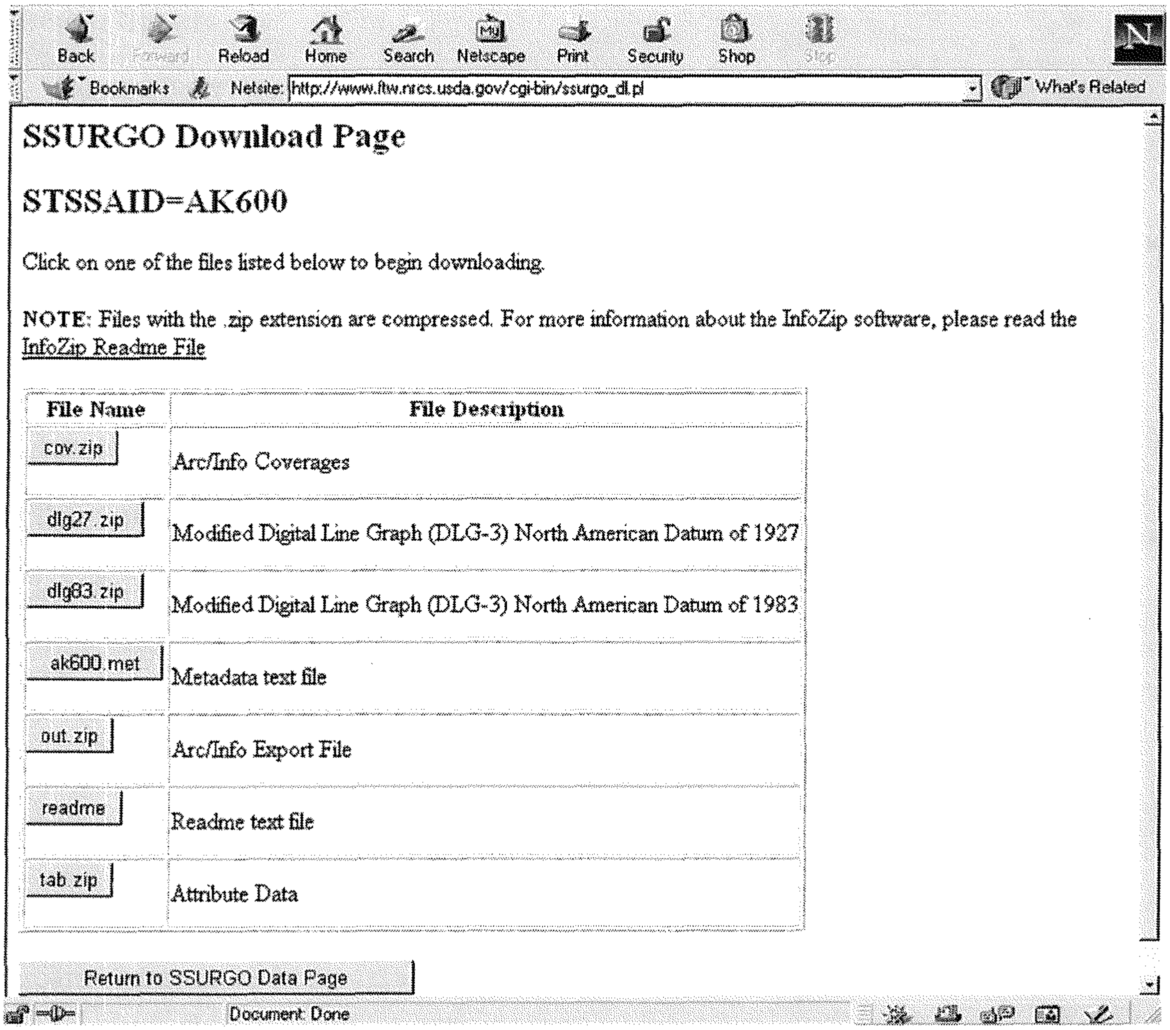


Figure 6. SSURGO Download page.

Figure 6 lists the files available within a certain SSURGO data set after the user clicks on one of the Download buttons on Figure 5. Clicking on any one of the "zip" file buttons under File Name starts the downloading of that file.

The software implementation described in this step simplified and automated the procedures to upload, validate, archive, update, and deliver digitized SSURGO data.

Step 3: Update

The SSURGO digitizing status map called for a Web technique for real-time updates that operates seamlessly on dynamic data (see Appendix C - Statement of Needs). Several Commercial Off-The-Shell (COTS) products and approaches to serving maps and processing updates were considered, including ESRI Spatial Database Engine (SDE), ESRI MapObjects, ESRI Internet Map Server (IMS), ESRI MapCafe, and Informix Dynamic Server Web Integration Option (DSWIO).

All of those COTS products were relatively new. They were too unstable, too slow, or too expensive to operate. SDE was platform-dependent; a certain version of SDE only worked with a certain version of Solaris OS, and a certain version of the Oracle RDBMS. It was difficult to optimize the engine; and nobody was ever sure of its performance. With MapObjects and IMS, server crashes were frequently encountered. The performance was also unacceptable for displaying maps. MapCafe required programming codes in the Java language. Bugs were discovered in the software while implementing some essential functions. Technical support and software revisions for MapCafe did not come timely to make any progress. Informix was being phased out by the department. It was also platform-dependent and not a feasible solution.

For those reasons, and after careful considerations of our design goals, a hybrid approach was selected. It was felt that relevant technologies evolve too rapidly to keep up. This approach, which is described in the next step, involved mixing and matching COTS and non-proprietary software components.

Step 4: Assessment

The first and hardest decision was on selecting a seamless method for online delivery of the SSURGO digitizing status map. The MIT Ortho Browser was chosen because it provides easy server-side management and thin-client usability access to digital orthophotos while freeing us from the external control and constraint of proprietary standards. This approach relied on the well-utilized CGI technology to build dynamic documents. The Perl scripting language was used for client-server communications. The C language was used to interface with various public domain graphics and shape libraries. The availability of these public utilities in the C language coupled with the transparent interface between them and the Perl code built for the MIT Ortho Browser made them a natural match.

The two primary libraries used in the prototype were:

- Vector Server for ESRI Shape Files (http://tull.mit.edu/orthoserver/vector) by John D. Evans (jdevans@mit.edu).
- ArcView Shape Library (http://gdal.velocet.ca/projects/shapelib) by Frank Warmerdam (warmerda@home.com).

The supporting libraries included:

- Cartographic Projections Library by Gerald Evenden (http://www.remotesensing.org/proj).
- Graphics Library for Fast Image Creation by Thomas Boutell (http://www.boutell.com/gd/).
- General Purpose Data Compression Library (http://www.cdrom.com/pub/infozip/zlib/).

On the COTS side, ESRI ArcInfo and ArcView GIS software was used in the initial stage to prepare the soils data files used for the map in the prototype. The newly agency-adopted RDBMS Oracle was used to store tabular information.

Step 5: Prototype

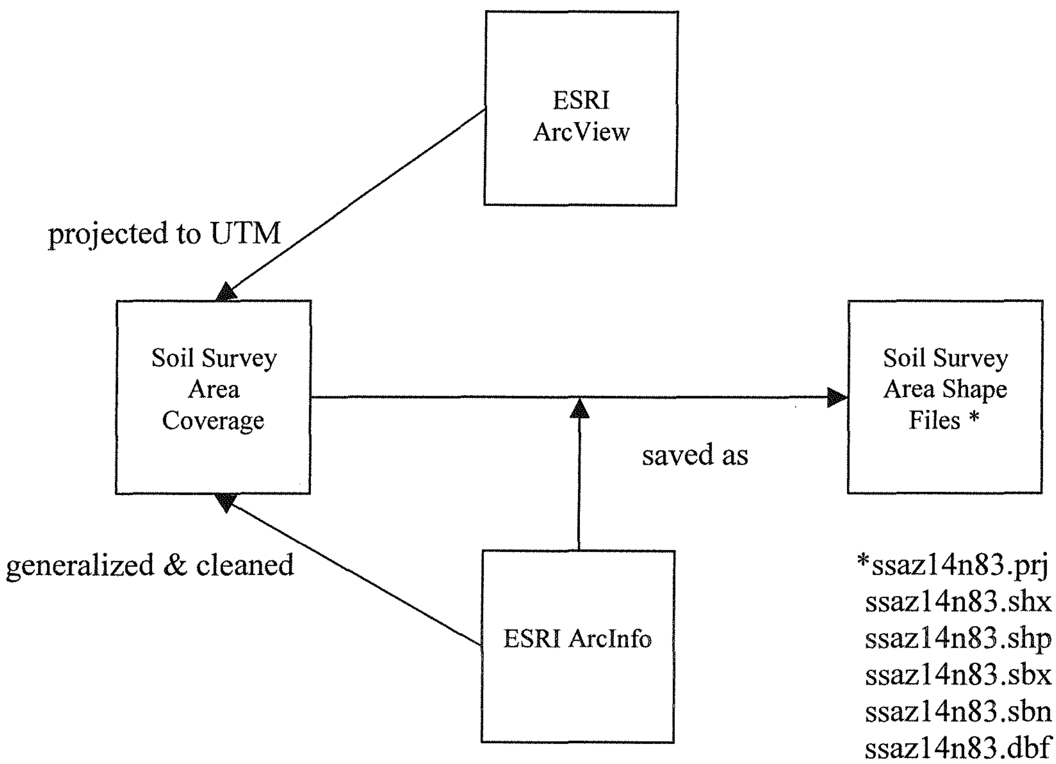


Figure 7. Projection and generalization.

Soil Survey Area coverage contains the boundaries information that was needed for displaying the map of SSURGO status data for the entire United States of America. The coverage was first re-projected from its Geographic format (LAT/LON) into UTM format (zone 14, NAD 83) using ESRI ArcView desktop GIS and mapping software. The coverage was then generalized and cleaned using ESRI ArcInfo GIS Workstation software to simplify and reduce its file size for quicker upload and display on the Web. The coverage was finally saved off as shape files and ready for the next step in the process.

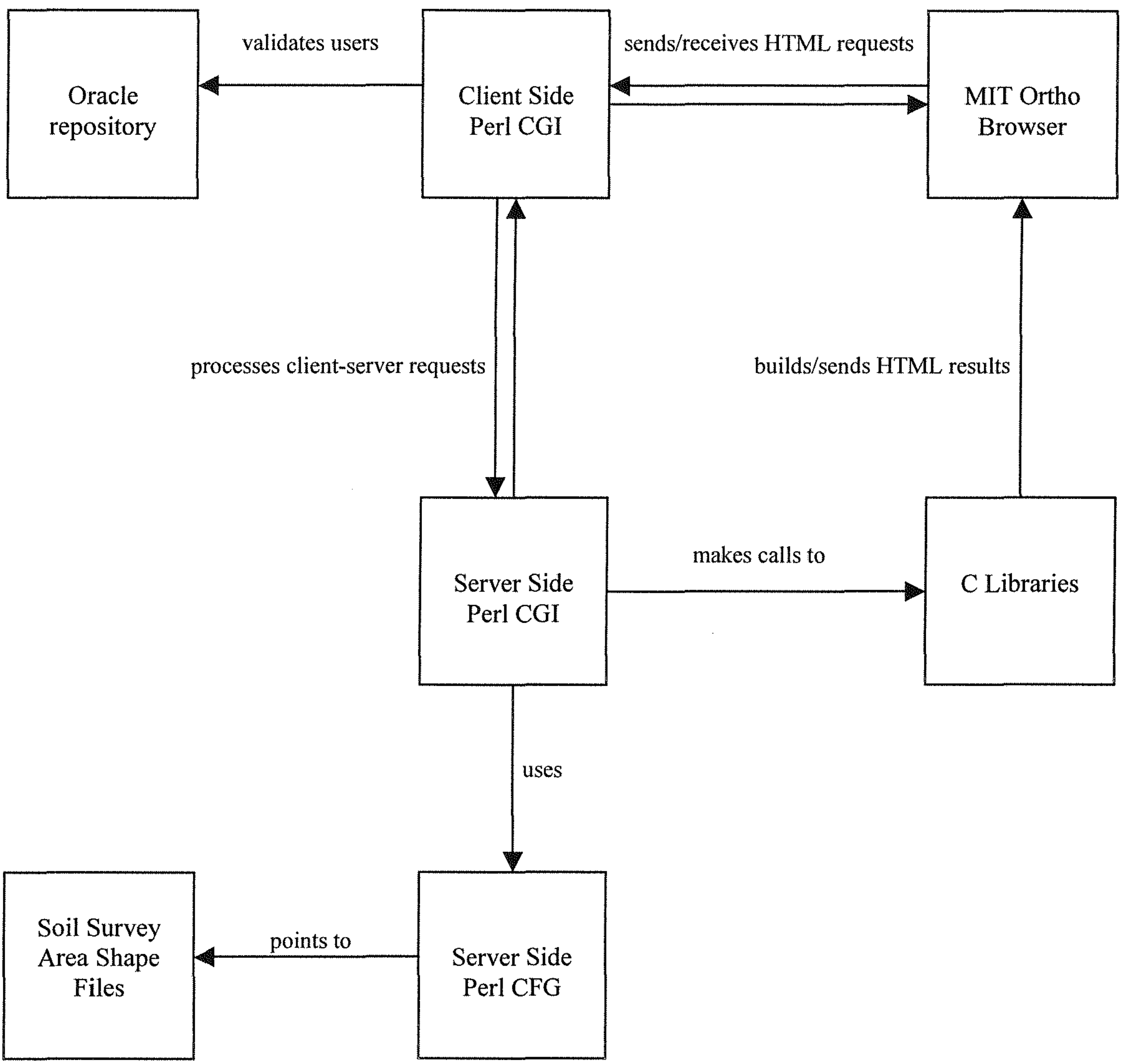


Figure 8. Prototype messaging.

The Perl code for the MIT Ortho Browser consists of three parts: a client side CGI code file, a server side CGI code file, and a configuration (CGF) code file. The client side CGI code sets up the image with user-selectable options, seamless image views, image lookups from a database, and seamless view port. It also handles all user actions coming from the browser. The server side CGI code reads and loads all settings and default values from the configuration (CFG) code file, performs image format conversions, and preprocesses images and shape layers before handing them over to the client CGI code. The configuration (CGF) code sets up image server options, and default values; and stores directory information on the locations of images and shape layers (including the location of the Soil Survey Area shape files).

The MIT Ortho Browser must run under Microsoft Internet Explorer (MSIE) due to its use of cascading style sheets. The prototype was developed and tested under MSIE, version 5.5. The following screen shots of the prototype will help to better explain the drawing in Figure 8 and the progressive steps taken to complete software implementation.

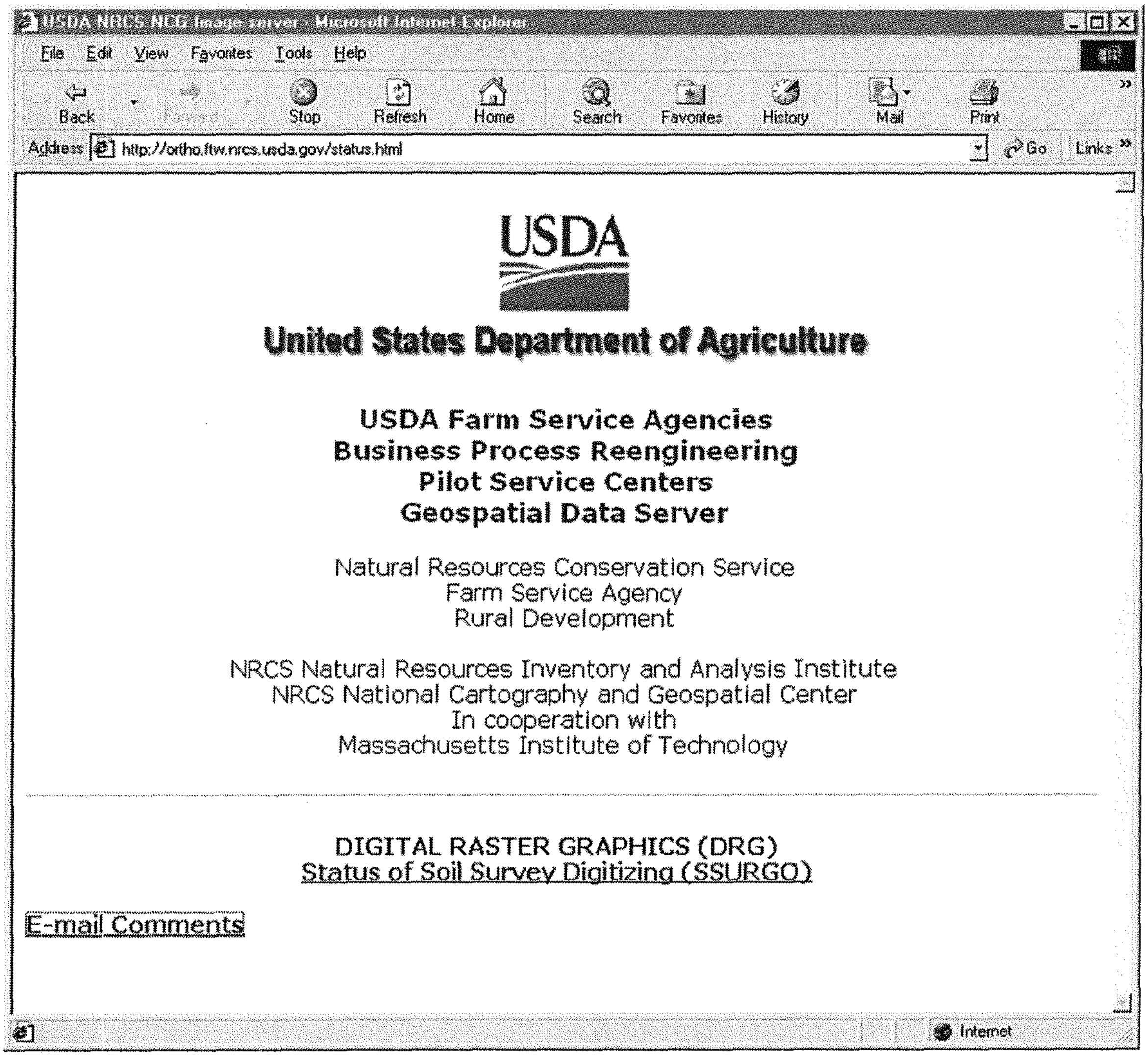


Figure 9. MIT Ortho Browser.

This screen provides the entrance to the MIT Ortho Browser. Clicking on the "Status of Soil Survey Digitizing (SSURGO)" starts up the client side CGI code (http://ortho.ftw.nrcs.usda.gov/cgi-bin/cdrgusa.cgi?datum=83&zone=14).

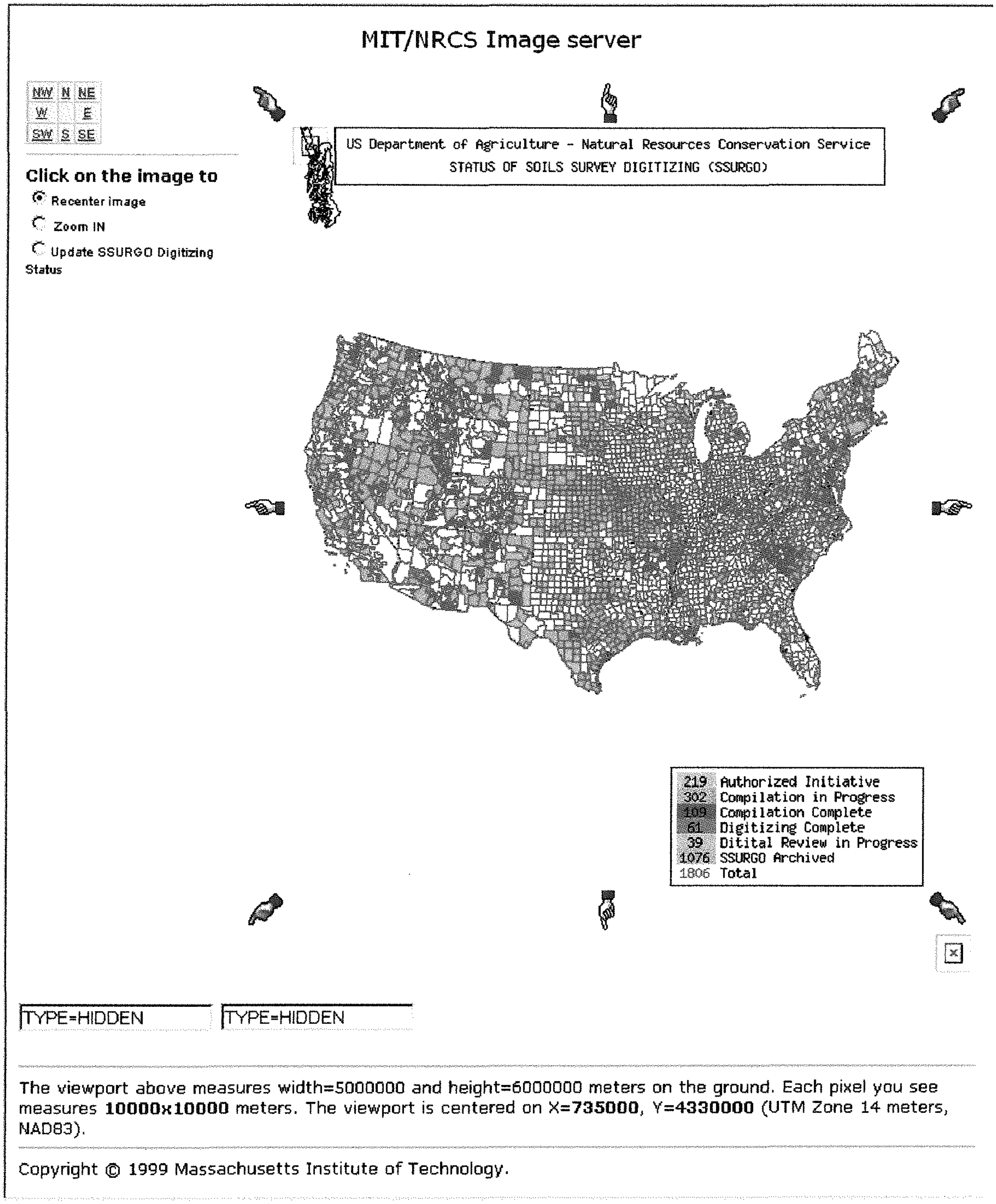


Figure 10. MIT/NRCS Image Server.

Figure 10 shows the main page displayed by the client side CGI code. A radio button was added to allow the updating of SSURGO status data. The server side CGI code makes calls to the Shapedraw utility from John Evans's Vector Server for ESRI Shape Files to draw the map. (Note: Several calls to Shapedraw are made per drawing or re-drawing of the map—one to draw the state lines, and one to draw to county lines.) Shapedraw was modified to use the latest versions of Thomas Boutell's Graphics Library for Fast Image Creation (gd-1.8.3), and Frank Warmerdam's ArcView Shape Library (shapelib-1.2.8). Calls to gdImageLine to draw polygons were replaced by calls to gdImagePolygon; and polygons were filled with calls to gdImageFilledPolygon. Re-projection code was also removed because it was not needed; and it would be too slow to re-project on the fly anyway. The map legend inside the graphics display area was also added to the Shapedraw utility. Agency standard thematic representations for SSURGO digitizing status are:

- Burleywood - (Authorized Initiative) - (RBG = 222, 184, 135)
- Light sky blue - (Compilation in Progress) - (RBG = 135, 206, 250)
- Cyan - (Compilation Complete) - (RBG = 0, 255, 255)
- Tomato - (Digitizing Complete) - (RBG = 255, 99, 71)
- Gold - (Digital Review in Progress) - (RBG = 255, 215, 0)
- Green Yellow - (SSURGO Archived) - (RBG = 173, 255, 47)

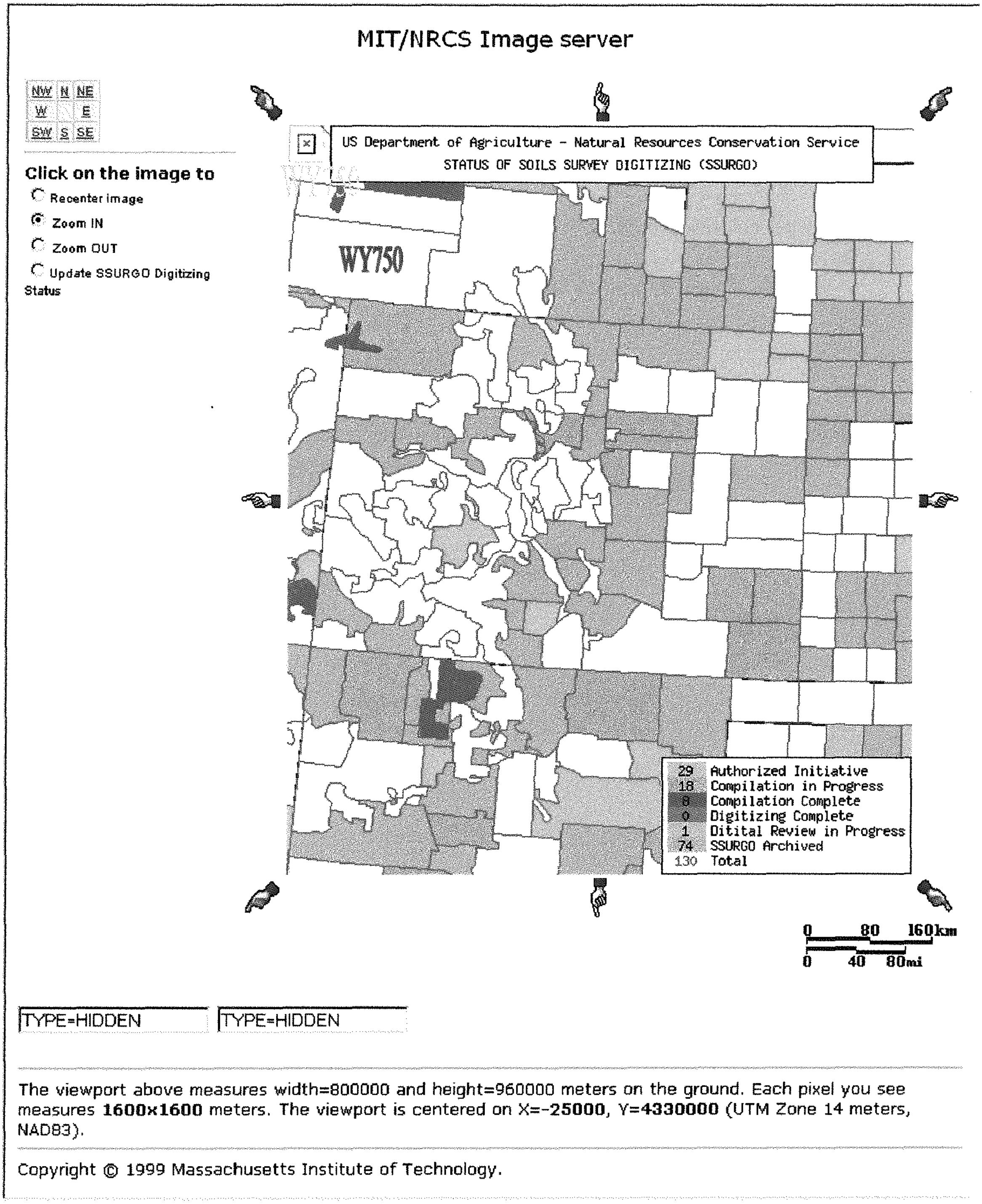


Figure 11. Zoomed-in image.

Figure 11 shows the zoomed-in image when the "Zoom In" radio button on Figure 10 was set, and a mouse click was applied in the proximity of the area whose status was to be changed. A message was sent from the MIT Ortho Browser to the client side CGI code—which in turn passed it to the server side CGI code and calling Shapedraw—to redisplay the image within the view centered on the mouse X and Y click.

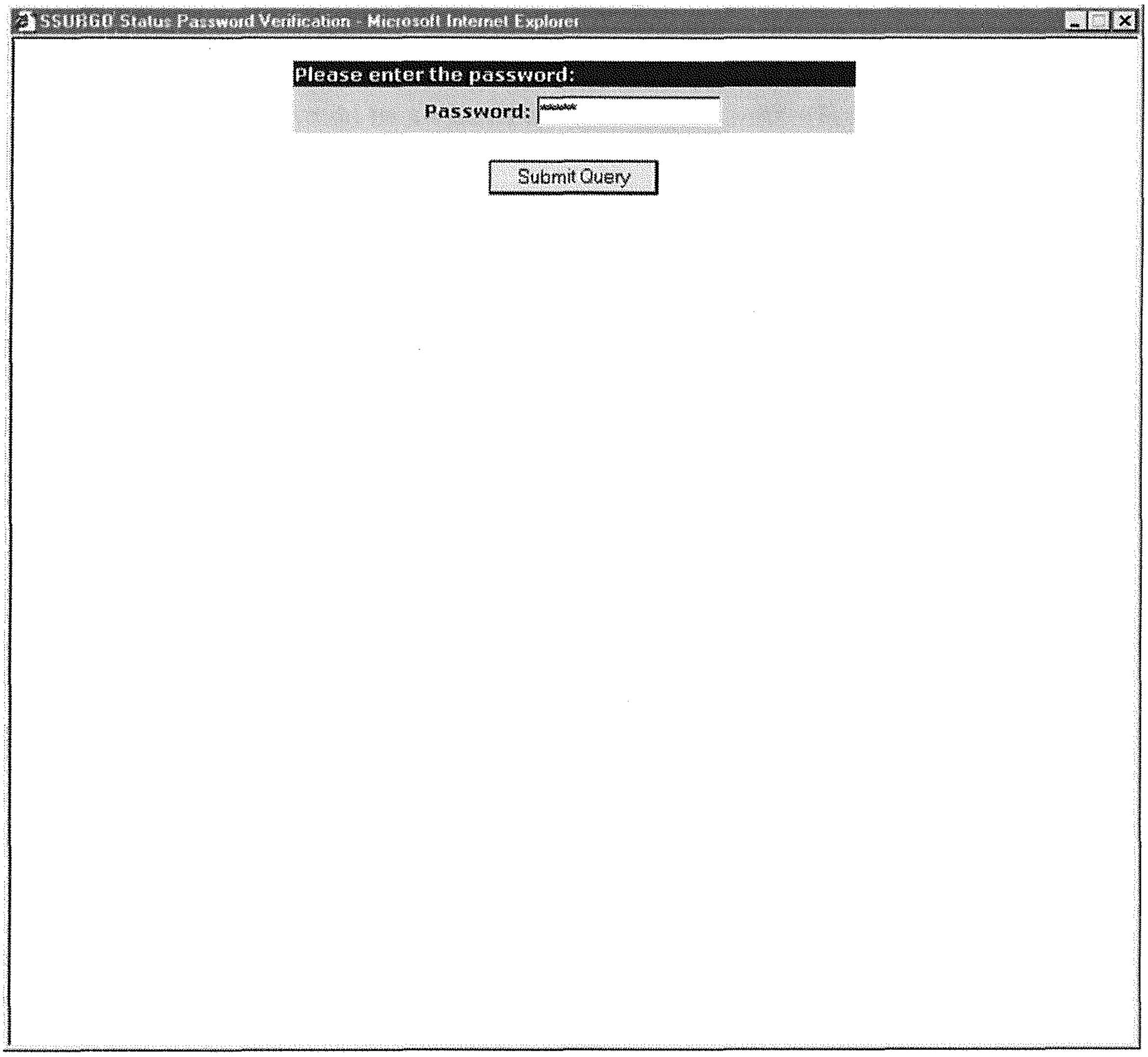


Figure 12. Password verification.

Figure 12 was displayed after the "Update SSURGO Digitizing Status" radio button on Figure 11 was set, and an area of the map was selected by the mouse. In this case, the second box from the top in the Northwest corner of the image, soil survey area WY750. (The Zoom In radio button got unset automatically because these buttons work exclusively of one another, thus allowing the system to process one screen command at a time.) A message was sent to the client side CGI code to process the command to update the selected SSURGO status. An HTML page was packaged and sent back to the Web as seen in Figure 12 to ask for a password. Authorized users who are allowed to update SSURGO status data would be aware of the correct password. (For the prototype, the password is "carto".)

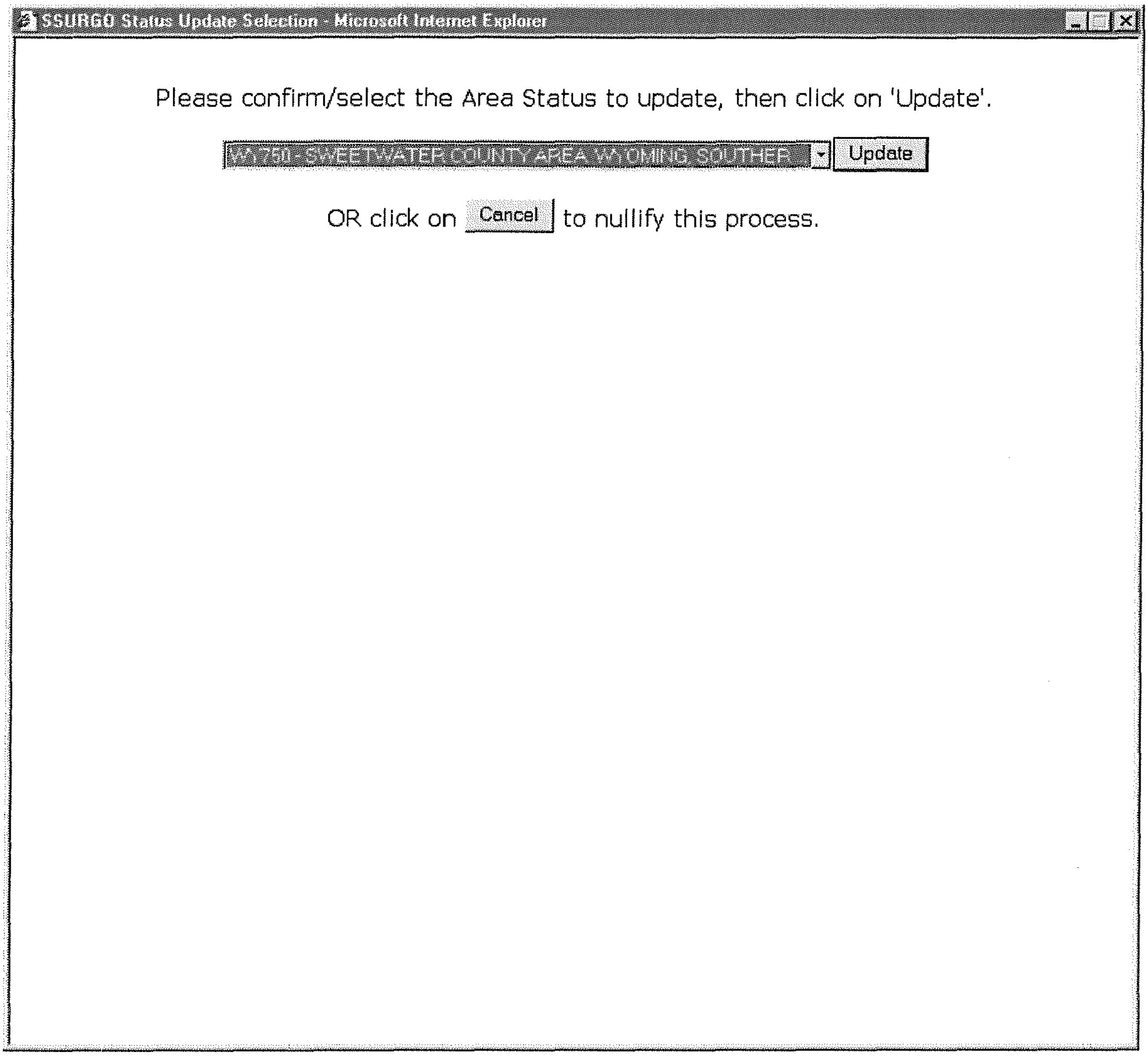


Figure 13. Confirm/Select.

After the password was entered and the user clicked on the "Submit" button, the password was sent back to the client side CGI code to be validated against the password stored in the Oracle repository. Once validated, a request was sent to the server side CGI code to read the selected status record from the Soil Survey Area shape files using the Attribute API of ArcView Shape Library. This record was then packaged in an HTML page and sent back to the Web as shown in Figure 13.

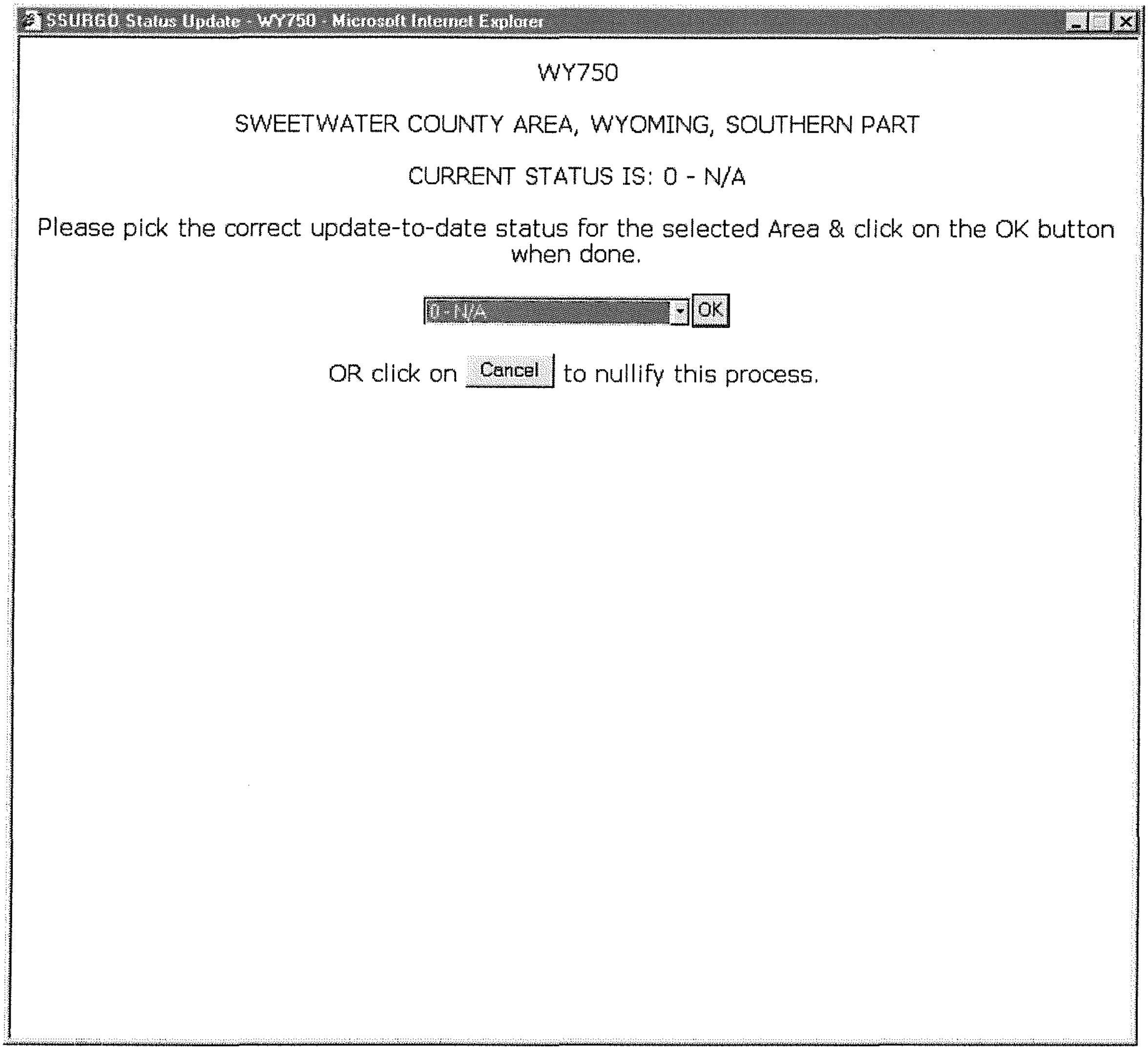


Figure 14. Update status.

Once the user confirmed the update on Figure 13, the complete record was displayed with soil survey area id (WY750), soil survey area name (SWEETWATER COUNTY AREA, WYOMING, SOUTHERN PART), and current SSURGO status (0 - N/A). A message was requested for the client side CGI code to package and send over to the Web an HTML page with the complete record information and a pull-down list of statuses.

The user then chose to update the current status of the selected SSURGO data set by selecting from the list, which includes the following items:

- 0 - N/A
- 1- Authorized Initiative
- 2 - Compilation in Progress
- 3 - Compilation Complete
- 4 - Digitizing Complete
- 5 - Digital Review in Progress
- 6 - SSURGO Archived

Each item in the list represents an integer value that corresponds to the value used for the Status attribute in the Soil Survey Area shape files.

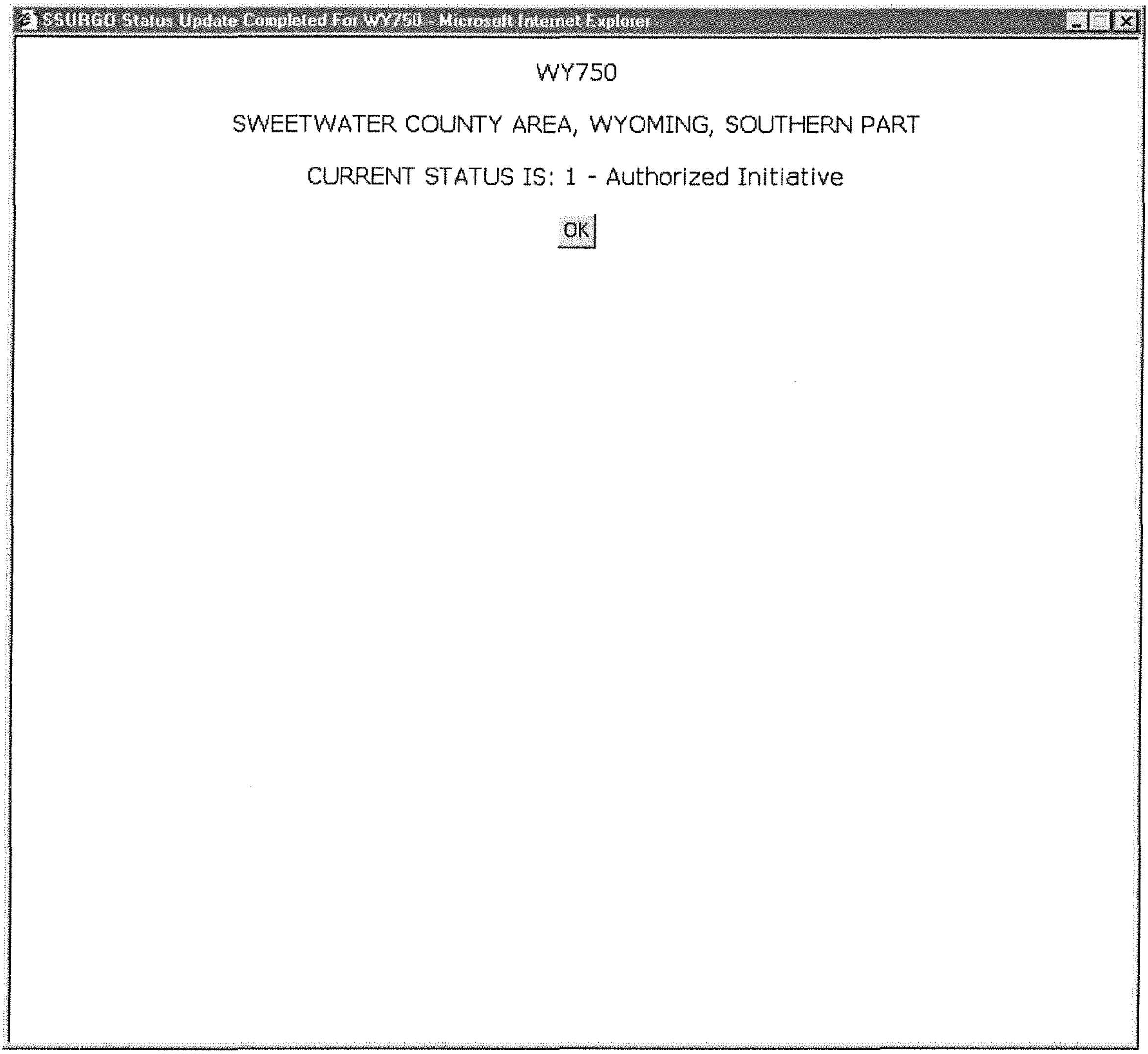


Figure 15. Update complete.

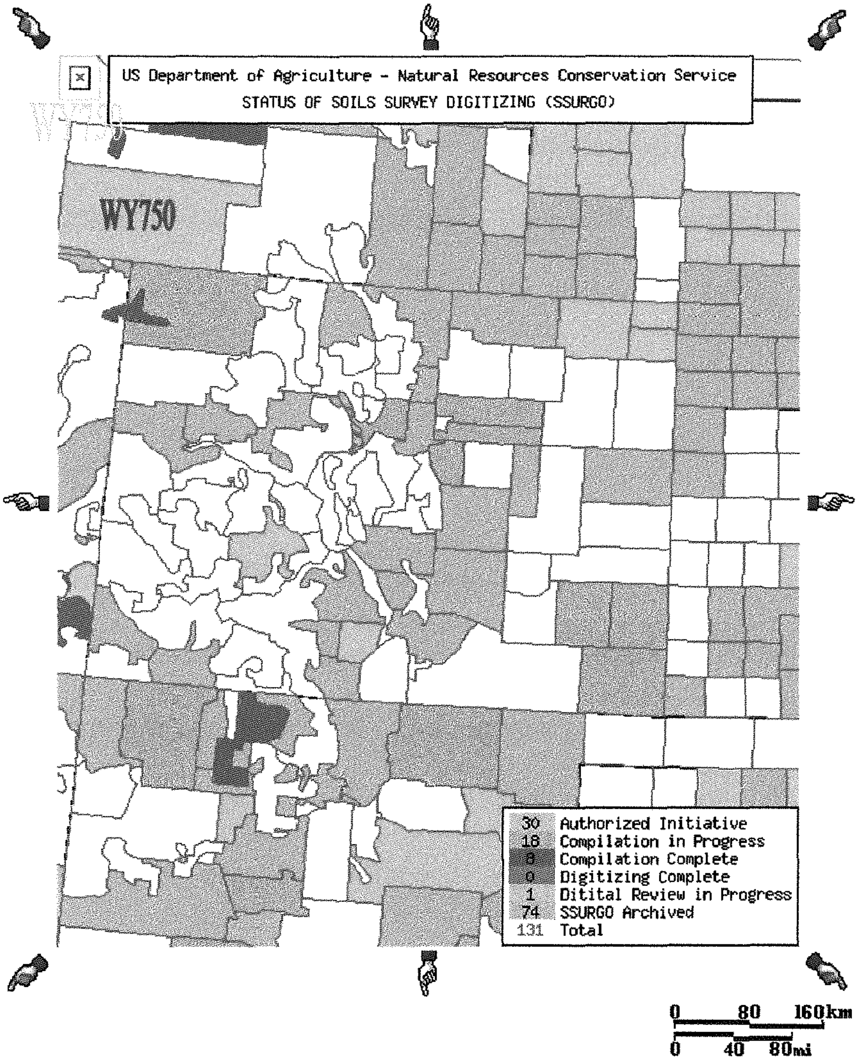
User selected item "1- Authorized Initiative" from the pull-down list on Figure 14 and clicked on "OK". The new status value was returned to the CGI codes where the Attribute API of the ArcView Shape Library was used to update the Status attribute of this SSURGO data set.

MIT/NRCS Image server

NW N NE
W E
SW S SE

Click on the image to

- Recenter image
- Zoom IN
- Zoom OUT
- Update SSURGO Digitizing Status



The viewport above measures width=800000 and height=960000 meters on the ground. Each pixel you see measures **1600x1600** meters. The viewport is centered on X=**-25000**, Y=**4330000** (UTM Zone 14 meters, NAD83).

Figure 16. Updated map.

Once the user clicked "OK" on Figure 15, notice that the status of soil survey area WY750 (the second box from the top in the Northwest corner of the image on Figure 11 and Figure 16) changed from white (status = 0 - N/A) to burleywood (status = 1 - Authorized Initiative). This change occurred in real time and was immediately available to everybody. The prototype as described above automates the SSURGO digitizing status map updates.

SUMMARY OF RESULTS

The software implementation for SSURGO data delivery and access simplified and automated the uploading, validation, archiving, and delivery of digitized SSURGO data. The procedures used in this process relied on UNIX shell scripts to query data for FTP posting and distribution. The Web database application offered a user-friendly interface for data delivery while reserving the capability to generate management accountability information.

The software implementation for the SSURGO digitizing status map prototype applied a dynamic Web technique to allow real-time updates of attributes data, and provide timely delivery of the map. A thin-client approach was used to shift the arduous map processing to the server. Non-proprietary technology was used to ease reliance on external control and constraint of standards. The use of cascading style sheet, which is a bandwidth-friendly technology, contributed to lessening the amount of time it takes to send the map over to the Web. Output from the MIT Ortho Browser log (see Appendix D - Computer Logs) showed it took from 1 to 6 seconds from inside the firewall to display the map. The average time was about 12 seconds from a laptop outside the firewall using a low speed 28K modem connected at the baud rate of 14400 bits per second.

The four design goals specified in the beginning have all been met. Web delivery was achieved and fulfilled the requirements for a consistent, ubiquitous, and platform-independent interface. Up-to-date information in the field was made available immediately for viewing, analyzing, or ad-hoc studies regardless of where the user is. Continuity and robustness were achieved with the procedure to update information in the field being independent of the individuals involved, and sufficiently accommodating to dissimilar users. This spatially enabling technology was implemented on existing software and hardware currently available at NCGC, proved to be reliable, performed satisfactorily, and fit in with NCGC future technology infrastructure including client access, bandwidth, distribution method, and support.

Chapter 5

Conclusions

Conclusions

The basic benefit of using dynamic update techniques for online maps and attributes data is the indispensable ability to obtain up-to-date information, and to handle persistent data changes and updates. It provides management with a much-desired mechanism to accountability and status information, and would save NRCS employees all over the country infinite hours of work to compile and to manually carry out progressive updates. The improved process would not only have a striking effect on the way workers perform their work; it would also save money and time. The results of the study supports the original hypothesis that it is feasible to deliver up-to-date online maps and attributes efficiently and effectively using dynamic update techniques.

Implications

The basic benefit of dynamic update techniques for online maps and attributes data is the direct savings in time and manpower. Teams of experts nationwide normally have to compile attributes information, make telephone calls, send emails, and wait for weeks before a static, digitized map can be produced and posted for viewing. The result still never reflects an up-to-date status. Indirectly, dynamic update techniques provide methods for documenting previously collected or produced geospatial data, and make that data documentation electronically accessible to everybody.

The result of this study has been highly anticipated. The prototype will be fully developed and put in service by NCGC. Its effect will be felt by the DUs nationwide. They will be able to efficiently perform status updates in real time with just a few clicks of the mouse. Managers and interested personnel will be delighted to have instant access to accurate and timely information.

Although the prototype and its performance clearly demonstrate the benefits of dynamic update techniques for online maps and attributes data, it took significant time and efforts to examine and study numerous relevant technologies. The newer and expensive technologies that were investigated had a tendency to confine themselves to a product line, were not easily applied in practical situations, or fell short of expectations in their performance. One of the reasons for this lack of success is the competitive nature of the business that makes it necessary for new products to be developed and marketed before they are ready.

The prototype developed in this study has an obvious limitation. The procedure to update attributes data currently allows a user to select any status value from the list displayed in figure 14. This could leave the database in a peril. Logically, a status graduates from "N/A" (i.e. value = 0) to "SSURGO Archived" (i.e. value = 6). Because a user is allowed to select any value from the list, it is possible that a status could go from "SSURGO Archived" (i.e. value = 6) to "Authorized Initiative" (i.e. value = 1). This status change does not make sense and put the integrity of the database in an undesirable situation.

Devising a technique to handle the problem presented in the above scenario is an important area for future research. The addition of expert knowledge to prevent users from assigning status values that do not make sense is a possible solution.

The dynamic update technique that was developed for SSURGO data in this study lays the groundwork for other important dynamic geospatial data that NCGC actively manages, such as STATSGO and HU. This work may also be extended to handle dynamic documents data. Examples of these data are technical notes, briefs, conservation practice standards, etc., which may be in ASCII, Microsoft Word, and Adobe Acrobat formats.

Recommendations

The nature of technology being constantly evolving necessitates that this study focused only on current and proven technologies that were relevant to the development and implementation of dynamic update techniques. As in any software development, it is expected that technical flaws may be discovered, and enhancements may be suggested with further testing and as the prototype evolves into a fully functional system. These issues and technology upgrades will be appropriately dealt with, evaluated, and selectively chosen for implementation.

It is suggested that the process described in this dissertation be continually evaluated and evolved once it is put to use. Emerging technologies affecting dynamic update techniques should be studied, and assessed against this process. Ensuing implementations should draw from, and improve upon, this work whenever possible. This dissertation should be electronically accessible and available to academia and the

U.S. government on UMI (University Microfilms International) Dissertation Services and government Web sites.

Summary

There was a need to accommodate perpetual data updates in the database and the delivery of real-time online maps and attributes information. Thus, the goal was set for this dissertation to analyze and evaluate dynamic update techniques and technology that would allow real-time delivery of up-to-date maps and attributes information. There were four specific design goals for the prototype system. They were: Web delivery; dynamic, mostly unattended upload; continuity and robustness; and implementation plan for NCGC.

A review of the literature was conducted. It revealed that in fact there was a need for Web-based updating systems and technology that allow efficient and timely delivery of information; and that dynamic update techniques were increasingly integrated in online systems that deliver maps and attributes data. It was also discovered that no research literature was available that specifically dealt with dynamic update techniques that provide for real-time delivery of up-to-date maps while allowing online update of attributes information.

The need was to integrate dynamic update techniques in a delivery system for online maps and attributes data. To achieve this integration goal, a five steps procedure was executed, culminating in the development of a prototype. Those steps included gathering data collection and storage information, investigating technological advances in data delivery and access, studying dynamic update techniques, accessing the feasibility of an implementation solution, and developing a prototype.

Business processes and conducts at NCGC were studied to clarify requirements and determine the focus for software implementation. The consensus was to simplify and automate the procedures to upload, validate, archive, update, and deliver SSURGO data sets; and to automate the SSURGO digitizing status map updates. Procedures were established to simplify the uploading, to coordinate the validation, and to automate the archive, update, and delivery of SSURGO data coming from DUs. These procedures relied on UNIX shell scripts to query data and other files for FTP posting and distribution. For the SSURGO digitizing status map, a dynamic Web technique was used to accomplish the updates.

The procedures for SSURGO data delivery and access depended on two primary UNIX shell scripts set up as CRON jobs. The first CRON runs a midnight everyday to pick up newly uploaded SSURGO data sets from the DU's private areas (Staging Area 1), validate the data, and move them to Staging Area 2. The second CRON runs at 2AM everyday and parses through Staging Area 2 looking for newly added and validated data sets. Additional machine processing is accomplished before these data sets can be archived and made available on the FTP site for public access. To facilitate delivery, a Web database application was developed that allows registered users to download SSURGO data sets via FTP. Customer download information is also kept in the database for management accountability purposes.

The procedures for real-time updates of SSURGO digitizing status map relied on a hybrid mix of COTS and non-proprietary software components, and a Web technique that operates seamlessly on dynamic data. The MIT Ortho Browser offered easy server-side management and thin-client usability; the CGI and Perl client-server strategy enabled document dynamics and communications; and the C language served as an interface with various public domain graphics and shape libraries.

The development of the prototype was based on the software foundation described above. First, the Soil Survey Area coverage containing status information was prepared by being re-projected using ESRI ArcView, then generalized and cleaned using ESRI ArcInfo. This coverage became the underlying database providing data for the status map. Next, the client, server, and configuration CGI scripts for the MIT Ortho Browser were tailored to display the Status of Soils Survey Digitizing (SSURGO) map and to handle updates. Significant changes and enhancements were made to the Shapedraw utility to manage filled polygons and map legends. Finally, additional HTML, Perl, and C codes were added to handle password processing, and to handle the individual soil survey area selections and updates.

In conclusion, this study adds to the pool of research literature currently available on the subject of dynamic update techniques. Its unique contribution is a dynamic update technique that provides real-time delivery of up-to-date maps while allowing online update of data attributes.

Appendix A

Correspondence

SDE Meeting Notes and Action Register

Updated November 13, 1998

Attending:

- Rob Vreeland
- Thanh Pham
- Su Liu
- Steve Nechero
- Dwain Daniels
- Fred Minzenmayer
- Hanh Nguyen
- Rosemary Rivera
- Paul Fukuhara
- Kim Prochnow
- Kristie McLeroy

Status to Date--

- All STATSGO spatial and attribute data is loaded in SDE/Oracle. SSURGO data for South Dakota is also loaded, this includes the spatial data and NASIS downloads of the attribute data. STATSGO is in NAD27 with geographic projections, SSURGO is in NAD83 also with geographic projections. **Action: Fred will grant permissions on the spatial data and Kristie will add logins for the additional persons (Su, Dwain, etc.) so that they can access the data.**
- General discussion as to whether all data should be NAD83. STATSGO will not be converted, but other datasets will be NAD83. Quads, etc. are already in NAD83, lat/long projection.
- General discussions of the table set up in Oracle, separate table spaces or whole? Appending to the table spaces may cause access problems due to the size of the tables. **Action: Set up a sub-committee to discuss, test, and decide issues of table access and loading of data into SDE/Oracle.**
- **Action: Thanh will load 3.? Update of SDE.**

Priorities—

- Status of Soil Surveys top priority. We need to have this preliminarily up and running by the 1st and 2nd weeks in December to demo to the NASIS meetings. Dwain will attend the meeting. **Action: Thanh will move spatial data to lat/long NAD83, Fred will acquire the NASIS download of the status of soil surveys, Rose will work with Thanh and Fred to automate the certified ssurgo listing that we keep here of available data.**
- 24K Quads loaded. **Action: Hanh, Paul, and Rob will load the Quarter Quad data and the USGS DOQ database. Paul will load ArcView 3.1.** Randy's county 100K data is already in lat/long projection so can be used as is. This data will be used to test ArcView applications with the ArcView IMS. **Action: Dwain and Su will test using ArcView.**

Major Software/Hardware Pieces and Responsible Parties— **Action: Each "primary" listed below will use this as an outline to establish work process steps under each element. This to be turned into Kim by Tuesday, November 17, 1998, noon.**

- SDE – running on Spark Station 3000 (iss55). Thanh is primary system a dministrator, Fred will serve as backup.
- Oracle – running on Spark Station 3000 (iss55). Kristie is primary system administrator, Hanh will serve as backup.
- ArcView IMS (UNIX) – running on Spark Station 3000 (iss55). Dwain will be primary. This software will be used to test the delivery of BPR data. Miami, KS was selected as the first test data to be loaded. **Action: Dwain and persons of his choice (suggested Elizabeth and Patsy) will load the Miami data.**
- Map Objects IMS (NT server) – running on Paul's machine at this time. Rob will be primary. **Action: Paul will initiate the purchase of a NT server to house this software. Software to be used for Arc Explorer projects, PRISM being the first. Rob will advertise the URL to use to test the PRISM data that is already loaded (Alabama data). Using Arc Explorer the URL is** http://paulnt.ftw.nrcs.usda.gov/scripts/esrimap.dll

Responses from the "primaries" will be consolidated into an active project plan to be used for training and funding planning purposes and as a timeline for activities. **Action: Kim will consolidate all responses from the "primaries" to have to Steve on Wednesday, November 18, 1998.**

Subject: **SSURGO data set successfully received**
Date: Wed, 2 May 2001 00:00:41 -0500 (CDT)
From: Super-User <root@ftw.nrcs.usda.gov>

```
This message was initiated by a batch job at NCGC, Ft. Worth, TX.

STSSAID=ny001

SSURGO data set ny001 was successfully moved out of the DU private
FTP location.

Thanks.
```

Subject: SSURGO data ny001 moved into FTP location
Date: Tue, 8 May 2001 02:14:15 -0500 (CDT)
From: Super-User <root@fgdc1.nrcs.usda.gov>
To: verify@ftw.nrcs.usda.gov, widu@ftw.nrcs.usda.gov

Subject: SSURGO data ny001 moved into FTP location

This message was initiated by a cron job at NCGC, Ft. Worth, TX.

STSSAID=ny001

This is to inform you that SSURGO data set ny001 has been moved to the SSURGO FTP location for online distribution. The metadata file has been parsed and is now available for the NSDI Clearinghouse search engine.

This message was sent by an automatic mailer, please DO NOT reply.

If you have any question or concern, please send an email to Thanh Pham at tpham@ftw.nrcs.usda.gov

Thanks.

Appendix B

Questionnaires

Data Collection & Storage questions presented to SMEs at NCGC

What kind of data do you serve?

Where and how do you get them?

How dynamic are they?

What do you have to do (e.g., preprocessing) before the data are ready to be served?

How do you serve the data (e.g., CD, FTP, Web, etc.)?

Appendix C

Validation Procedures

```
NCGC Procedures for Verifying Certified SSURGO Data via FTP
(procedures and checklist may vary for off-site archiving)

1) Ensure accessability of the following required Items:
       SCS-CGI-019 requesting archival
       Certification Letter
       Index of Quadrangles in survey
       data set of survey being reviewed
       Waiver of Scale if other than 1:12 or 1:24K

2) Load Data
       _cd to /gdb/ssurgo/online99/*du to check for new ftp data
       _cd to major working partition, usually /data/ssurgo,
        and place "unzip" script here
       _cp -r /gdb/ssurgo/online99/*du/stssaid .
       _cd to stssaid and check for missing directories and files
       _unzip all then remove *.zip
       _verify that that following directories and files exit
               _cov
               _dlg (If the Horizontal Datum Name is NAD83,or)
               _dlg27 (and)
               _dlg83 (If the Horizontal Datum Name is NAD27)
               _*.met
               _out
               _readme
               _tab

3) Print out portions of the data as follows (ex. for FQ NAD27)
       cd cov
       ls -la | lp (pipe a long listing to your printer destinat.)
       cd ../dlg27
       head -20 s*.0af | lp
       head -20 s*.0sf | lp
       cd ../dlg83
       head -3 s*.0af | lp
       head -3 s*.0sf | lp
       cd ..
       lp *.met
       cd out
       ls -ls | lp

   For FQ NAD83
       cd cov
       ls -la | lp
       cd ../dlg
       head -20 s*.0af (s*.*af for QQ)
       head -20 s*.0sf (s*.*sf for QQ)
```

```
        cd ..
        lp *.met
        cd out
        ls -la | lp
```

Use /data/ssurgo/SSURGO_ARCHIVING_FTP/checklist and printouts
to verify using check-mark or X to indicate if the directories
have/do not have the following:

```
cov     _ stssaid_a
        _ stssaid_b
        _ stssaid_l (if survey includes linear special features)
        _ stssaid_p (if survey includes spec. soil pt. or adhoc feat.)
        _ stssaid_q
        _ info
        _ readme
        _ required coverages (s......._0a, _0p, _01)
dlg (or)
dlg27 (and)
dlg83
        _readme
doc     _metadata
out     _ stssaid_a.e00
        _ stssaid_b.e00
        _ stssaid_l.e00
        _ stssaid_p.e00
        _ stssaid_q.e00
        _ comp.e00
        _ compyld.e00
        _ feature.e00
        _ forest.e00
        _ helclass.e00
        _ hydcomp.e00
        _ inclusn.e00
        _ interp.e00
        _ layer.e00
        _ mapunit.e00
        _ mucoacre.e00
        _ muident.e00
        _ muyld.e00
        _ plantcom.e00
        _ plantnm.e00
        _ rangenm.e00
        _ readme
        _ rsprod.e00
        _ s......._a.e00, _l.e00, _p.e00
        _ ssacoac.e00
        _ ssarea.e00
        _ taxclass.e00
        _ windbrk.e00
        _ wlhabit.e00
        _ woodland.e00
        _ woodmgt.e00
        _ yldunits.e00
```

Readme -readable and complete

```
_ cd  /job_loc/county_st/cov
```

_ Check for any log or scratch (xxxs.....) files in cov dir

_ Compare the cov directory print-out with that of out dir. Insure that for every coverage there exists a corresponding file in the out directory, and that for every stssaid_a, _b, _l, _p and _q there exists and _a.e00, _b.e00, _l.e00, _p.e00 and _q.e00. The cov directory should also contain a readable readme file and an info directory.

_Compare the contents of the out directory with the tab directory and the Map Interpretations Record data base section of the metadata to insure that they all have the same information regarding the relational tables.

_ The following tables of the tab directory must be populated:
- _ codes (must be at least 2669 lines)
- _ comp
- _ hydcomp
- _ interp
- _ layer
- _ mapunit
- _ mucoacre
- _ plantnm
- _ ssacoac
- _ ssarea (check ssarea name and eddate format)
- _ taxclass
- _ yldunits

These additional tables may or may not be populated:
- _ compyld
- _ feature (only if survey has special soil or ad hoc features)
- _ forest
- _ helclass (if populated, must have more that 2 fields- see instructions below to "fix" if improperly populated)
- _ inclusn
- _ muident
- _ muyld
- _ plantcom
- _ rangenm
- _ rsprod
- _ windbrk
- _ wlhabit
- _ woodland

The following instructions are provided to "fix" the helclass table if it is improperly populated:
- _ cd /data/ssurgo/stssaid/tab
- _ vi helclass
- _ delete all but 1st two lines, wq! (ex: if helclass table consists of 45 lines, arrow down 2 lines then type 43dd) save and quit (esc, wq!)
- _ cd ../cov
- _ invoke arc
- _ Arc: info
- _ ENTER USER NAME> ARC
- _ ENTER COMMAND> SEL HELCLASS
- _ ENTER COMMAND> PURGE
- _ THIS COMMAND WILL DELETE SELECTED RECORDS. OK?> Y

```
_ ENTER COMMAND> Q STOP
_ Arc: rm ../out/helclass.e00
_ ARC> export info helclass ../out/helclass.e00
_ ARC> q
_ % pwd (should still be /data/ssurgo/stssaid/cov)
_ % more log
_ % rm log
_ %  cd ../out
    _ ls -la helclass.e00 - (should have block size of 650)
_ cd ../tab
   _ ls -la helclass - (should have block size of 85)
   _"more" helclass - (make sure it looks unpopulated)
_ cd ../doc _ vi met*
   _ add "Table not populated" to helclass in the MUIR
     information portion of the Entity and Att. Info. Sec.
```

4a. On-screen verification for dlg accuracy

```
_ cd dlg27
        ls s*.0af | wc -l (generates count quickly)
        ls s*.0aa | wc -l
        ls s*.0sf | wc -l
        ls s*.0sa | wc -l
_ cd ../dlg83
        ls s*.0af | wc -l
        ls s*.0aa | wc -l
        ls s*.0sf | wc -l
        ls s*.0sa | wc -l
```

_ Ensure that af count = aa count and sf count - sa count.
(note: dlg27 count may be unequal to dlg83 count if the limit of survey coincides with the quadrangle limit. This is a good starting point to verify that no "empty" dlg's exist due to a loss of data in the projection process from NAD27 to NAD83. You should suspect an "empty" dlg if there are only 4 line records-neatlines- in the soils dlgs. To confirm, cd to corresponding s*******.0aa. If there are only two line records, for UNIV and BLANK, you can be relatively certain that an "empty" dlg exits. Document, then delete the empty s........0af and s........0aa. You will notice that there is no corresponding coverage in the cov directory.)

4) Using dlg header printouts:
- _ check s #'s (compare soils and special feature dlgs, compare with s #'s listed in USGS Topo. Names Base section in metadata
  - _ check if ESRI Version 05-08-92
- _ Check quadrangle names against USGS index and against the USGS Topographic Names Base Section of the Metadata for spelling errors and accuracy, and make sure format is quadrangle name, state abbreviation
  - _ check date and revision date
  - _ check scale(A waiver of scale is required if scale is something other than 12000 or 24000)
  - _ USDA/NRCS SSURGO DATA
  - _ If Datum listed is NAD27, check for Datum Flag of "0"
  - _ If Datum listed is NAD83, check for Datum Flag of "1"

- _ Insure that scale listed is consistant with resolution
- _ View coordinates - pay particular attention is survey is in more that one zone. Compare the coordinates in the first few node records to insure that they are within the limits of the stated zone coordinates
- _ Check for the underscore between "SPECIAL_FEATURES"

5) Review metadata file using /data/ssurgo/SSURGO_ARCHIVING_FTP /meta.template.eval of numbered elements and pay particular attention to items with "*" which indicate elements where most errors repeatedly occur.

6) Check tabular data online, especially "codes", "feature", "ssarea" and "helclass"

7) Errors noted? Document and notify responsible party(s)

8) Data Clean? - continue with archival process
Instructions to make iso file and cut cd's:

```
create directory under /data/ssurgo for iso_usr (ex iso_co)
put a copy of the readme text file here.
copy data from data ssurgo to iso_user
ex: cd /data/ssurgo/iso_co
    cp -r /data/ssurgo/ks085 .
    ls -la (should have readme and ks085)
```

9) cd /data/ssurgo/iso_cdrom and make ks085.iso as follows:
mkisofs -o ks085.iso -r -V ks085 /data/ssurgo/iso_co

10) Using the established "shortcuts" to FTP and Easy_CD_Pro dbl. click FTP to open session and login to host machine using user name and password. Hit "open"
Change local drive from C:\ to D:\
open iso_cdrom and put files on local drive.

From the main menu bar, select "window", and pulldown to choose "new_command_line_window" which will open a new window with a unix prompt FTP(sss18)>
At the prompt type: cd /data/ssurgo/iso_cdrom
the return line should say: CWD /data/ssurgo/iso_cdrom
followed by another prompt FTP(sss18)> type "binary"
the return line should read Binary connection complete
with another prompt at which type "ls -la"
you should see your iso file here and another prompt
FTP(sss18)> type "mget ks085.iso"
when all files are transferred you will get a return prompt
and you are ready to use the shortcut to "Easy_CD_Pro" to
find your data on drive D and cut the cd's ....easy!

Check the cd's, complete a media library form and take to Geospatial Databases "sto" & "osp" numbers and archiving

11) notify responsible party to post data to ftp site

12) update digstats table in "lib" database

```
/data/ssurgo/SSURGO_ARCHIVING_FTP/meta.template.eval

Evaluation of Numbered Elements in Metadata Template

Use print-out of metadata file to edit. Make approved edit
corrections and document all.
* denotes elements which are commonly incorrectly populated

Citation_Information:
     (1) Publication_Date: ___(1)___ date all data are completed
         enter as 8 digits in following format: YYYYMMDD
         example: 19990204
     (2) Title: Soil Survey Geographic (SSURGO) database for
         ___(2)___
        * Title should be consistant with the Classification and
        Correlation Document and ssarea name in the ssarea table of
        muir, as well as template elements (20), (24) and title
        in the 1st Process Step if survey was previously published,
        (45), and (46).
     (2a) stssaid in lower case

Description:
     (3) format - 3.75 or 7.5 minute quadrangle format

Time_Period_of_Content:
     (4) Beginning_Date: is the same as ___(1)___
     (4a) Ending_Date: revision date (if the data are revised) or
          the same as ___(1)___ and ___(4)___ if data not revised

Spatial_Domain:
        (5) West_Bounding_Coordinate: *always "-" in US
        (6) East_Bounding_Coordinate: *always "-" in US
        (7) North_Bounding_Coordinate:
        (8) South_Bounding_Coordinate:
        * carry 3 decimal places for 1:24000 and
                4 decimal places for 1:12000
        Compare the coordinates listed with the max-min limits
        as shown on the indexplotpg which was generated during
        processing of the SSURGO Evaluation amls.

Place_Keywords:
        (9) State name
        (10) County name
             * use separate line for each county
        (11) USGS Topographic Map Names Data Base
        (11) Name of quadrangle                 (s.......)
             * reference USGS Topographic Quadrangle Index and GNIS
Names Base
             * be sure to include the word "Quadrangle" after
quadrangle names
             and parenthesis ( ) around s numbers
             * Make sure these names coincide with soils and special
feature
             dlg header names and s numbers.

Point_of_Contact:
```

```
   Contact_Information:
      Contact_Address:
         (12) address
         (13) city
         (14) State or Province
         (15) Postal Code
         (16) Voice Telephone
              Contact_TDD/TTY_Telephone: 202 720 7808
         (17) Fax Telephone
         * make sure current and consistant with (53)-(58)
         * use hyphens ONLY in (15) Postal Code

Cross_Reference:
   Citation_Information:
         (18) Originator:  Natural Resources Conservation Service,
__(18)__
         * Soil Conservation Service if prior to 1994
         (19) Publication Date
         * use calendar year survey was published for previously
         published data, or "unpublished material" if not published
         (20) Title of Survey
         * should be consistant with other references to published
         survey, including, but not limited to (2), (24), 1st Process
         Step, (45) and (46)
Data_Quality_Information:
   Logical_Consistancy_Report:
         (20a) "joins note" information
         * joins information should be placed at end of this paragraph,
         not added as a process step

Data_Quality_Information:
   Attribute_Accuracy:
         (21) accuracy in acres
         * should be 5 - 6 for 1:24000, can be as low as 2.5 - 3 for
         1:12000 for order 2

Lineage:
   Source_Information:
      Source_Citation:
          (22) Originator
          (23) Publication Date - Calendar Year *
          (24) Title *
          (25) Geospatial Presentation Form *
          (26) Publication Place
          (27) Publisher
          (28) Source Scale Denominator
          (29) Type of Source Media *
          (30) Source Time Period (Single Date/Time or Range)
          (31) Source Currentness Reference
          (32) Source Citation Abbreviation
          (33) Source Contribution

* Make sure that Cross Reference elements (18), (19) and (20)
are consistant with Lineage entries.

Process_Step
         (34) Process Description
```

* if survey is listed in cross reference, it should be reflected in Source Citation and 1st Process Step, which should include title, date and scale of published survey as well as an evaluation statement of whether or not the soil delineations and map unit components were found to be accurate or obsolete.

A process step should be included which reflects the current survey including compilation, digitizing and/or scanning, software used, quality control and other pertinent information.

A process step should be included referencing the Map Unit Interpretations Record Data Base developed by NRCS Soil Scientists according to national standards

(36) Source Citation

(35) Process Date

```
Spatial_Data_Organization_Information:
  Direct_Spatial_Reference_Method: Vector
Spatial_Reference_Information:
  Horizontal_Coordinate_System_Definition:
    Planar:
      Grid_Coordinate_System:
        Grid_Coordinate_System_Name: Universal Transverse Mercator
        (37) UTM Zone Number
             * use separate entry for each zone
        (38) Longitude of Central Meridian
             * use separate line for each zone, always "-" in US,
             carry to one decimal place    ex: -85.0
     Planar_Coordinate_Representation: *
        * although separate lines should be used for each zone in the
        grid coordinate system section, only one entry is needed for
        the planar coordinate representaion section, even if there are
        more than one zones in the survey
        (39) Abcissa Resolution *0.305 for 1:12000, 0.61 for 1:24000
        (40) Ordinate Resolution 0.305 for 1:12000, 0.61 for 1:24000
Geodetic_Model:
        (41) Horizontal Datum Name * North American Datum of 1927
        or North American Datum of 1983
        (42) Ellipsoid Name * Clarke-NOT Clark-1866 or GRS80
        (43) Semi-major Axis * 6378206.4 or 6378137.0
        (44) Denominator of Flattening Ratio - 294.98 or 298.257

Entity_and_Attribute_Information:
  Detailed_Description:
    Attribute:
      Attribute_Domain_Values:
        Codeset_Domain:
          Codeset_Name:
            Classification and Correlation of the Soils of
_____(45)_____
          Codeset_Source:
            U.S. Department of Agriculture, Natural Resources
            Conservation Service
            * Codeset Name should be consistant with  all other title
entries
Overview_Description:
  Entity_and_Attribute_Overview:
    Map Unit Delineations are described by the Map Unit Interpretations
      Record database. This attribute database gives the proportionate
```

extent of the component soils and the properties for each soil. The database contains both estimated and measured data on the physical and chemical soil properties and soil interpretations for engineering, water management, recreation, agronomic, woodland, range, and wildlife uses of the soil. The soil Map Unit Interpretations Record database consists of the following relational tables:

codes (database codes) - stores information on all codes used in the database
comp (map unit component) - stores information for soil map unit components
compyld (component crop yield) - stores crop yield information for soil map unit components
forest (forest understory) - stores information for plant cover as forest understory for soil map unit components
helclass (highly erodible lands class) - stores the highly erodible land classification for wind and water assigned to the soil map units
hydcomp (hydric component information) - stores data related to the hydric classification, criteria, landform, etc.
inclusn (map unit inclusion) - stores the names of soils included in the soil map units
interp (interpretation) - stores soil interpretation ratings (both limitation ratings and suitability ratings) for soil map unit components
layer (soil layer) - stores characteristics of soil layers for soil map unit components
mapunit (map unit) - stores information that applies to all components of a soil map unit
mucoacre (map unit county acres) - stores the number of acres for the map unit within a county
muyld (map unit yield) - stores crop yield information for the soil map unit
plantcom (plant composition) - stores plant symbols and percent of plant composition associated with components of a soil map unit
plantnm (plant name) - stores the common and scientific names for plants used in the database
rangenm (range name) - stores the range site names
rsprod (range site production) - stores range site production information for soil map unit components
ssacoac (soil survey area county acreage) - stores the acreage for the county within the boundary of the soil survey area
ssarea (soil survey area) - stores information that will apply to an entire soil survey area
taxclass (taxonomic classification) - stores the taxonomic classification for soils in the database
windbrk (windbreak) - stores information on recommended windbreak plants for soil map unit components
wlhabit (wildlife habitat) - stores wildlife habitat information for soil map unit components
woodland (woodland) - stores information on common indicator trees for soil map unit components
woodmgt (woodland management) - stores woodland management

```
        information for soil map unit components
      yldunits (yield units) - stores crop names and the units used to
        measure yield
      * "Table not populated" should be used where appropriate

      Special features are described in the feature table.  It includes
a
      feature label, feature name, and feature definition for each
      special and ad hoc feature in the survey area.
      * Paragraph above should be deleted if no special or ad hoc
feature
      have been included in the survey

Distribution_Information:
  Distributor:
    Contact_Information:
      Contact_Organization_Primary:
        Contact_Organization: U.S. Department of Agriculture, Natural
                              Resources Conservation Service, National
                              Cartography and Geospatial Center
      Contact_Address:
        Address_Type: mailing address
        Address: P.O. Box 6567
        City: Fort Worth
        State_or_Province: Texas
        Postal_Code: 76115
      Contact_Voice_Telephone: 800 672 5559
      Contact_Facsimile_Telephone: 817 509 3469
  Resource_Description: ______(46)______ SSURGO
        * consistant with all title entries
        * do not include the words "soil survey" in the Resource Desc.
Standard Order Process
  Digital_Form:
      Digital_Transfer_Information:
        Format_Name: ARC/INFO coverage
        Format_Information_Content: spatial
        Transfer_Size: ___(48)___
        * du -sk . example: 20700 = 20.7
      Digital_Transfer_Option:
        Offline_Option:
          Offline_Media: CD-ROM
          Recording_Format: ISO 9660 Level 1
    Digital_Form:
      Digital_Transfer_Information:
        Format_Name: ARCE
        Format_Information_Content: spatial
        Transfer_Size: ___(48)___
      Digital_Transfer_Option:
        Offline_Option:
          Offline_Media: CD-ROM
          Recording_Format: ISO 9660 Level 1
    Digital_Form:
      Digital_Transfer_Information:
        Format_Name: DLG
        Format_Version_Date: ___(47)___
        * currently 19920508
        Format_Specification: Optional
```

```
        Format_Information_Content: spatial and keys
        Transfer_Size: ___(48)___
      Digital_Transfer_Option:
        Offline_Option:
          Offline_Media: CD-ROM
          Recording_Format: ISO 9660 Level 1
    Digital_Form:
      Digital_Transfer_Information:
        Format_Name: ASCII
        Format_Information_Content: keys and attributes
        Transfer_Size: ___(49)___
      Digital_Transfer_Option:
        Offline_Option:
          Offline_Media: CD-ROM
          Recording_Format: ISO 9660 Level 1
     Fees:
       The charge is $50 for a CD-ROM that contains one or more data
      sets. A data set is one soil survey area in full quadrangle
      format and includes both spatial and attribute data.
        (50) format Full Quadrangle or Quarter Quadrangle
Metadata_Reference_Information:
  Metadata_Date: ___(51)___
  * Date first created, this should also be changed if data are revised
  Metadata_Contact:
    Contact_Information:
      Contact_Organization_Primary:
        Contact_Organization: U.S. Department of Agriculture, Natural
                              Resources  Conservation Service
      Contact_Position: State Soil Scientist
      Contact_Address:
        Address_Type: mailing address
        Address: ___________(53)___________
        City: ___________(54)___________
        State_or_Province: __(55)__
        Postal_Code: __(56)__
      Contact_Voice_Telephone: ___________(57)___________
      Contact_Facsimile_Telephone: ___________(58)___________
      * must be consistant with contact information in "Point of
        Contact" section, hyphens used only in Postal Code
  Metadata_Standard_Name: Content Standard for Digital Geospatial
Metadata
  Metadata_Standard_Version: FGDC-STD-001-1998
```

List of Errors or Omissions Critical Enough to Warrant Correction

During the trial period for transfer of SSURGO data via ftp, questions arose from the NCGC SSURGO Review Team as to what errors would be critical enough to keep the data from being posted for online distribution.
What follows is a list, and brief explanation, of the review process for the metadata elements.

1. The Publication_Date must by in YYYYMMDD format. This format must be used for compliance with Metadata Standards Version 2.
2. The Title must be the official name of the survey area. The official name is usually found in the Classification and Correlation Document.

3. The quadrangle format must be filled in (either 3.75 or 7.5) and match the 8.3 file names. If the format is 3.75' then the file names should have a 1, 2, 3, or 4 suffix. If the format is 7.5' the suffix should be 0.
4. The Time_Period_of_Content must be in YYYYMMDD format. This format must be used for compliance with Metadata Standards Version 2.
5. The Bounding_Coordinates must reflect the complete area that the survey covers. These coordinates let the customer know exactly where the survey is located and how much area is included.
6. The State and County names must be correct in the Place_Keyword element. This lets the customer know if the survey area covers more than one state or county.
7. The quadrangle names in the USGS Topographic Map Names Data Base must be complete. This is only place that all the quadrangles are listed for the survey area. It must be correct. If there are 30 quadrangles in the survey area, they all should be listed here and be correctly spelled. The 8.3 filenames listed after the quadrangle names must be correct.
8. A minimum size delineation must be included. Example: usually six acres for a scale of 24,000.
9. In the Lineage, the quadrangle format (quarter or full) must match the spatial data. Also, the scale (12,000 or 24,000) must match the spatial data.
10. The UTM_Zone_Number and the Longitude_of_Central_Meridian must match the spatial data. If the utm zone in the DLGs is 14, then the zone in the metadata should also be 14.
11. The Datum information must match the spatial data. If the datum information in the DLG is NAD27, then the metadata should also reflect this.
12. The Resource_Description under Distributor should match, exactly, the Title. These elements refer to the same data set and should be the same.
13. All four Transfer_Sizes should be filled in. This lets the customer know the size of the data set that will be downloaded.
14. The Metadata_Date must be in the format YYYYMMDD. This format must be used for compliance with Metadata Standards Version 2.

Statement of Needs

- Users shall be able to dynamically upload maps and attributes data (i.e. geospatial and tabular data) from anywhere in the field.
- Up-to-date maps and attributes data shall be uploaded to a central database.
- Up-to-date maps and attributes data shall be stored and managed in a spatial-enabled Relational Database Management System.
- Users shall be able to view up-to-date maps.
- Users shall be able to view up-to-date attributes information.
- Authorized users shall be able to perform real-time update of maps and attributes information.
- The procedure to update information in the field shall accommodate dissimilar users.
- The system interface shall be platform-independent.
- The system interface shall be consistent across all users.
- Maps and attributes data shall be displayed to users in an accurate and timely fashion.

Appendix D

Computer Logs

Output from the MIT Ortho Browser log:

```
START: Year=2001, Month=5, Day=23, Hour=16, Minute=3, Second=2
END:   Year=2001, Month=5, Day=23, Hour=16, Minute=3, Second=3

START: Year=2001, Month=5, Day=23, Hour=16, Minute=3, Second=2
END:   Year=2001, Month=5, Day=23, Hour=16, Minute=3, Second=8

START: Year=2001, Month=5, Day=23, Hour=16, Minute=6, Second=3
END:   Year=2001, Month=5, Day=23, Hour=16, Minute=6, Second=3

START: Year=2001, Month=5, Day=23, Hour=16, Minute=6, Second=3
END:   Year=2001, Month=5, Day=23, Hour=16, Minute=6, Second=4

START: Year=2001, Month=5, Day=23, Hour=16, Minute=16, Second=44
END:   Year=2001, Month=5, Day=23, Hour=16, Minute=16, Second=45

START: Year=2001, Month=5, Day=23, Hour=16, Minute=16, Second=44
END:   Year=2001, Month=5, Day=23, Hour=16, Minute=16, Second=50

START: Year=2001, Month=5, Day=24, Hour=10, Minute=46, Second=38
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=46, Second=39

START: Year=2001, Month=5, Day=24, Hour=10, Minute=46, Second=38
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=46, Second=43

START: Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=32
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=32

START: Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=32
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=33

START: Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=56
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=56

START: Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=56
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=47, Second=57

START: Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=20
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=20

START: Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=20
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=21

START: Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=26
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=26
```

```
START: Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=26
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=27

START: Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=53
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=53

START: Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=53
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=48, Second=54

START: Year=2001, Month=5, Day=24, Hour=10, Minute=49, Second=11
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=49, Second=11

START: Year=2001, Month=5, Day=24, Hour=10, Minute=49, Second=11
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=49, Second=12

START: Year=2001, Month=5, Day=24, Hour=10, Minute=50, Second=29
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=50, Second=29

START: Year=2001, Month=5, Day=24, Hour=10, Minute=50, Second=29
END:   Year=2001, Month=5, Day=24, Hour=10, Minute=50, Second=30

START: Year=2001, Month=5, Day=24, Hour=11, Minute=3, Second=41
END:   Year=2001, Month=5, Day=24, Hour=11, Minute=3, Second=41

START: Year=2001, Month=5, Day=24, Hour=11, Minute=3, Second=40
END:   Year=2001, Month=5, Day=24, Hour=11, Minute=3, Second=47

START: Year=2001, Month=5, Day=24, Hour=11, Minute=36, Second=2
END:   Year=2001, Month=5, Day=24, Hour=11, Minute=36, Second=2

START: Year=2001, Month=5, Day=24, Hour=11, Minute=36, Second=10
END:   Year=2001, Month=5, Day=24, Hour=11, Minute=36, Second=16

START: Year=2001, Month=5, Day=24, Hour=11, Minute=37, Second=4
END:   Year=2001, Month=5, Day=24, Hour=11, Minute=37, Second=5

START: Year=2001, Month=5, Day=24, Hour=11, Minute=37, Second=11
END:   Year=2001, Month=5, Day=24, Hour=11, Minute=37, Second=13
```

Reference List

Abiteboul, S. (1999). On views and XML. In *Proceedings of the Eighteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (pp. 1-9). New York: ACM Press.

Acharya, A., & Badrinath, B. R. (1996, October). A framework for delivering multicast message in networks with mobile hosts. *ACM Mobile Networks and Applications,* 1(2), 199-219.

Acquah, J. B. H. (1991). Imagys: A language for interactive spatial modeling and graphics (programming language). *Dissertation Abstract International*, 52, AAT 9115721.

Adam, N., Yesha, Y., Awerbuch, B., Bennet, K., Blaustein, B., Brodsky, A., Chen, R., Dogramaci, O., Grossman, B., Holowczak, R., Johnson, J., Kalpakis, K., McCollum, C., Neches, A.-L., Neches, B., Rosenthal, A., Slonim, J., Wactlar, H., Wolfson, O., & Yesha, Y. (1996, December). Strategic directions in electronic commerce and digital libraries: towards a digital agora. *ACM Computing Surveys,* 28(4), 818-835.

Andresen, D., Yang, T., Ibarra, O. H., & Egecioglu, O. (1998, February). Adaptive Partitioning and scheduling for enhancing WWW application performance. *Journal of Parallel and Distributed Computing,* 49(1), 57-85.

Balasubramanian, V., & Bashian, A. (1998). Document management and Web technologies: Alice marries the mad hatter. *Communications of the ACM,* 41(7), 107-115.

Balasubramanian, V., Bashian, A., & Porcher, D. (1997). A large-scale hypermedia application using document management and Web technologies. In *Proceedings of the Eighth ACM Conference on Hypertext* (pp. 134-145). New York: ACM Press.

Bezanson, W. R. (1995). Performance support: online, integrated documentation and training. *Proceedings of the 13th Annual ACM Conference on Systems Documentation, USA,* (1995, September), 1-10.

Broadbent, M., Butler, C., Hansell, A, & Dampney, C.N.G. (1995, February). Business value, quality and partnerships: Australasian information systems management issues. *Australian Computer Journal,* 27(1), 17-26.

Bromberek, J., & Medina, R. (1998). *Functional assessment of Enigma INSIGHT 4.1.* Chicago, IL: Doculabs.

Brown, I. M. (1999). Developing a virtual reality user interface (VRUI) for geographic information retrieval on the Internet. *Transactions in GIS,* 3(3), 207-220.

Bryron, S., Kenwright, D., Cox, M., Ellsworth, D., & Haimes, R. (1999, August). Visually exploring gigabyte data sets in real time. *Communications of the ACM,* 42(8), 83-90.

Caputi, M. J. (1998, November). Developing real-time digital audio effects for electric guitar in an introductory digital signal processing class [CD-ROM]. *IEEE Transactions on Education,* 41(4), 341-341.

Castro, E. (1996). *HTML for the World Wide Web.* Berkeley, CA: Peachpit Press.

Coad, P., & Yourdon, E. (1990). *Object-oriented analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Corn, M. (1999, February). Digital libraries: The view from National Library of Medicine. *IEEE Computer,* 32(2), 47-47.

Cornell, G., & Horstmann, C. S. (1996). *Core Java.* Mountain View, CA: SunSoft Press.

Croft, W. B., & Savino, P. (1998, January). Implementing ranking strategies using text signatures. *ACM Transactions on Information Systems,* 6(1), 42-62.

Danesh, A. (1997). *JavaScript interactive course.* Corte Madera, CA: The Waite Group.

Deep, J., & Holfelder, P. (1996). *Developing CGI applications with Perl.* New York: John Wiley & Sons.

Entlich, R., Olsen, J., Garson, L., Lesk, M., Normore, L., & Weibel, S. (1997, April). Making a digital library: the contents of the CORE project. *ACM Transactions on Information Systems,* 15(2), 103-123.

Elmasri, R., & Navathe, S. B. (1994). *Fundamentals of database systems* (2nd ed.). Redwood City, CA: The Benjamin/Cummings Publishing Company.

Entlich, R., Olsen, J., Garson, L., Lesk, M., Normore, L., & Weibel, S. (1997, April). Making a digital library: the contents of the CORE project. *ACM Transactions on Information Systems,* 15(2), 103-123.

Environmental Systems Research Institute. (1998). *GIS by ESRI: ArcExplorer.* Redlands, CA: Author.

Evans, J. (1997). *Infrastructures for sharing geographic information among* *environmental agencies*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Massachusetts.

Faundeen, F., & Zanter, K. (1997, September). *Metadata management, integration, and visualization through the U.S. geological survey's global land information system*. Poster session presented at the Second IEEE Metadata Conference, Silver Spring, MD.

Faloutsos, C. (1985, March). Access methods for text. *ACM Computing Surveys,* 17(1), 49-74.

Folsche, W. R. (1996). Letter from the director. *National Cartography and Geospatial Center annual report* (pp. 2-3). Fort Worth, TX: USDA - Natural Resources Conservation Service.

Fry, M. R. (1996, August). Management of the IT industry: A survey of critical challenges. *Australian Computer Journal,* 28(3), 88-95.

Geppert, L. (1998, June). IC design on the world wide web. *IEEE Spectrum,* 35(6).

Grare, L. (1995, January/February). Making a case for 12-in. optical discs. *IMC Journal,* 31(1), 10-11.

Griffin, S. (1999, February). Digital libraries: The view from National Science Foundation. *IEEE Computer,* 32(2), 46-46.

Harder, C. (1998). *Serving maps on the Internet: Geographic information on the World Wide Web*. Redlands, CA: Environmental Systems Research Institute.

Hoff, A. V., Shaio, S., & Starbuck, O. (1996). *Hooked on Java.* Reading, MA: Addison-Wesley Developers Press.

Horvath, E. (1996, September). *The National Cartography and Geospatial Center, Geospatial Databases branch, 1997 business plan*. (Available from National Cartography and Geospatial Center, 501 Felix St., Bldg. 23, Fort Worth, Texas 76115)

Horstmann, C. S. (1997). *Practical object-oriented development in C++ and Java.* New York: John Wiley & Sons.

Huang, Y.-W., & Yu, P. S. (1998, May). A bandwidth-sensitive update scheduling method for Internet push. In *Proceedings of the 18th International Conference on Distributed Computing Systems* (pp. 303-310). Los Alamitos, CA: IEEE Computer Society.

Jones, S. B. (1994). Caught in the World Wide Web: MIT moves computer documentation online. In *Proceedings of the XXII ACM SIGUCCS Conference on User Services* (pp. 187-190). New York: ACM Press.

Kamel, N., & King, R. (1992, December). Intelligent database caching through the use of page-answers and page-traces. *ACM Transactions on Database Systems,* 17(4), 601-646.

Kent, P., & Kent, J. (1997). *Official Netscape JavaScript 1.2 book* (2nd ed.). Research Triangle Park: Ventana Communications Group.

Kirstein, P., & Montasser-Kohsari, G. (1996, June). The C-ODA project: Online access to electronic journals. *Communications of the ACM,* 39(6), 87-99.

Litwin, W., Mark, L., & Roussopoulos, N. (1990, September). Interoperability of multiple autonomous databases. *ACM Computing Surveys,* 22(3), 267-293.

Lizard Tech, Inc. (1999). *MrSID Wavelet Based Image Compression.* Retrieved February 17, 2000, from the World Wide Web: http://www.lizardtech.com/products/mrsid/mrsid_whatis/.

Lucent Technologies - Lucent NetworkCare. (2000). *Lucent NetworkCare - Performance Management Software Solutions.* Retrieved February 17, 2000, from the World Wide Web: http://www.ins.com/software/.

Mackert, L., & Lohman, G. (1989, September). Index scans using a finite LRU buffer: A validated I/O model. *ACM Transactions on Database Systems,* 14(3), 401-424.

Mauch, J. E., & Birch, J. W. (1993). *Guide to the Successful Thesis and Dissertation: A Handbook for Students and Faculty* (3rd ed.). New York: Marcel Dekker.

McKelvy, M., Spotts, J., & Siler, B. (1997). *Using Visual Basic 5* (2nd ed.) [Special Edition]. Indianapolis: Que.

Nambisan, S., & Wang, Y.-M. (1999, January). Roadblocks to Web technology adoption. *Communications of the ACM,* 42(1), 98-101.

Neerincx, M. A., Kerstholt, J. H., Dijk, E.M.A.G., Nijholt, A., Pemberton, S., & Spenkelink, G.P.J. (Eds.) (1999). *Web-based services for information and commerce: User-centered design principles, methods, and applications.* (Available from Telematics Institute, Drienerlolaan 5, 7522 NB Enschede)

Newton, P. W., Taylor, K. L., Trinidad, G. S., Ackland, R. G., & Abel, D. J. (1999). GIS networks and the marketplace. *Transactions in GIS,* 3(2), 91-95.

Orendorf, J., & Kacmar, C. (1996). A spatial approach to organizing and locating digital libraries and their content. In *Proceedings of the 1st ACM International Conference on Digital Libraries* (pp. 83-89). New York: ACM Press.

Orland, B., Wu, X., & Chavan, A. (1997). Internet and Intranet: Technology adoption in natural resources management. In *Proceedings of ASPRS/ACSM/ RT'97* (vol. 4, pp. 265-276). Washington, DC: American Society for Photogrammetry and Remote Sensing.

Padula, M., & Rinaldi, G. R. (1999, July-August). Mission-critical web applications: a seismological case. *Interactions,* 6(4), 52-66.

Plaisant, C., Jain, V. (1994). Dynamaps: dynamic queries on a health statistics atlas. In *Proceedings of the CHI '94 Conference Companion on Human Factors in Computing Systems* (pp. 439-440). New York: ACM Press.

Plasil, F., Balek, D., & Janecek, R. (1998, May). SOFA/DCUP: Architecture for component trading and dynamic updating. In *Proceedings of the International Conference on Configurable Distributed Systems* (pp. 43-52). Los Alamitos, CA: IEEE Computer Society.

Raphaely, D. (1997). *Oracle8 application developer's guide* (part no. A58241-01). Redwood Shores, CA: Oracle Corporation.

Roehle, B. (1997, March). Channeling the data flood. *IEEE Spectrum,* 34(3), 32-38.

Rosencrantz, M. J. (1997, November). Push technology. *The CPA Journal*, 11, 34-39.

Rowson, J. (1998). DAC '98 panelist bios and position statements. In P. Miller (Ed.), *IEEE Spectrum, Special Features, Design Automation Conference 1998*. Retrieved February 17, 2000, from the World Wide Web: http://www.spectrum.ieee.org/special/dac98/rowson.html

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. (1991). *Object-oriented modeling and design.* Englewood Cliffs, NJ: Prentice Hall.

Sauder, D. A., & Lubell, J. (1999). The express web server: a user interface for standards development. In *Proceedings of the 1999 Symposium on Applied Computing* (pp. 610-615). New York: ACM Press.

Schatz, B., & Chen, H. (1999, February). Digital libraries: Technological advances and social impacts. *IEEE Computer,* 32(2), 45-50.

Schatz, B., Mischo, W., Cole, T., Bishop, A., Harum, S., Johnson, E., Neumann, L., Chen, H., & Ng, D. (1999, February). Federated search of scientific literature. *IEEE Computer,* 32(2), 51-59.

Stallings, W., & Slyke, R. V. (1998). *Business data communications* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Stiles, R., Tewari, S., & Mehta, M. (1997). Adapting VRML 2.0 for immersive use. In *Proceedings of the 2nd Symposium on Virtual Reality Modeling Language* (pp. 75-81). New York: ACM Press.

Swartz, R. (1990). *UNIX applications programming. Mastering the Shell.* Carmel, IN: SAMS.

Thomas, S. W., Alexander, K., & Guthrie, K. (1999, February). Technology choices for the JSTOR online archive. *IEEE Computer,* 32(2), 60-65.

Tyner, J. (1992). *Introduction to thematic cartography.* Englewood Cliffs, NJ: Prentice-Hall.

University Consortium for Geographic Information Science. (1996). Research priorities for geographic information science. *Cartography and Geographic Information Systems,* 23(3), 115–127.

USDA - Natural Resources Conservation Service. (1994a, April). *National handbook of conservation practices.* Retrieved February 17, 2000, from the World Wide Web: http://www.ftw.nrcs.usda.gov/nhcp_2.html.

USDA - Natural Resources Conservation Service. (1994b, April). *NRCS home page.* Retrieved February 17, 2000, from the World Wide Web: http://www.nrcs.usda.gov/.

USDA - Natural Resources Conservation Service. (1995, November). *Data rich and information poor.* Washington, DC: Author.

USDA - Natural Resources Conservation Service. (1996, October). *Strategic plan for fiscal years 1996-2002.* Washington DC: Author.

USDA - Natural Resources Conservation Service. (1998, August). *Soil survey schedule NASIS 4.0 - User manual.* Washington DC: Author.

USDA - Service Center. (1996, December). *America's private land, a geography of hope.* Washington DC: Author.

USDA - Service Center. (1997, August). *Future directions: A vision of information technology for field conservationists.* Washington DC: Author.

USDA - Service Center. (1998, August). *Geographic information system (GIS) strategy.* Washington DC: Author.

Watters, C., & Shepherd, M. (1990, April). A transient hypergraph-based model for data access. *ACM Transactions on Information Systems,* 8(2), 77-102.

Whitten, J. L., Bentley, L. D., & Barlow, V. M. (1994). *Systems analysis and design*

Methods (3rd ed.). Burr Ridge, IL: Irwin.

Wilensky, R. (1996, May). Toward work-centered digital information services. *IEEE Computer,* 29(5), 37-44.

Wortman, K. (1994, July). Developing standards for a national spatial data infrastructure. *Cartography and Geographic Information Systems*, 21(3), 145-149.