A Prototype for Enterprise GIS

by

Marc S. Witkowski, Paul M. Rich, and Gordon N. Keating

GISLab, EES-9, Earth and Environmental Sciences Division, MS-D452, Los Alamos National Laboratory, Los Alamos, NM 87544

Abstract

An increasing number of institutions are challenged with implementing robust geographic information system (GIS) capabilities for a large number of individuals through information-sharing and interconnected networks. In the past, numerous technological roadblocks hampered the successful implementation of enterprise GIS (EGIS). With the advent of high-speed networks; increasingly fast computers; intelligent, spatial-data serving technologies; improved data architecture; and advances in GIS software; the newest challenge involves integration of the various technological and institutional components. This integration can be viewed as an inevitable stage in the evolution of GIS. We propose an EGIS design based on three elements: 1) distinct stakeholder roles, 2) a complete geospatial data cycle, and 3) proven data warehouse concepts. Four groups of stakeholders play key roles: data providers, data managers, GIS users, and customers. A complete geospatial data cycle ensures that all aspects of data and workflow are implemented and integrated. Spatial data warehouse design guides the flow of work and data through three steps: staging, storage, and delivery. We detail a five-step methodology for implementation of EGIS: 1) design specification, 2) data resource evaluation, 3) spatial data warehouse design, 4) physical system architecture design, and 5) implementation. This five-step process has been effective in the implementation of the Cerro Grande Rehabilitation Project (CGRP) GIS. The goal of the CGRP-GIS is to provide EGIS capability in response to the May 2000 Cerro Grande wildfire, which forced evacuation of the town of Los Alamos and Los

Alamos National Laboratory. The CGRP GIS technical design uses the spatial data warehouse design concepts and custom enterprise tools for data and work tracking, together with a data policy and standard procedures of data and work flow, to achieve a complete geospatial data cycle. The result is an EGIS design with broad applicability.

TABLE OF CONTENTS

1 Introduction

An enterprise geographic information system (EGIS) provides access to shared geospatial information and analysis resources for a large number of concurrent users located in different parts of an institution (Rich et al. 2001, Somers 2002) (Figure 1). At its core, this GIS capability typically includes hardware, software, and shared geospatial data, accessed from centralized servers through high-speed intranet or Internet networks. In addition to the hardware, software, and data, there are other important, often-overlooked components of technical design that are essential for successful implementation of EGIS, involving in particular process integration and distinct roles of participants (e.g., Oppmann 1999). Thus the EGIS can also be defined as an *effort* to design integrated geospatial management techniques to serve a complex institution (Peng et al. 1998). Depending on the needs of an institution, EGIS can be designed to emphasize the full spectrum of GIS capabilities (Figure 1). Although EGIS is receiving increasing attention, there has been no comprehensive consideration of technical design issues.

The primary function of any information system is the collection, storage, and analysis of data to improve the decision-making process (Calkins and Tomlinson 1977). The benefits of GIS for data institution and analysis in a map-based context have long been exploited for large geographic information management efforts, such as continent-wide census projects (e.g., CGIS in Canada and TIGER in the U.S.). The advent of fast, powerful PCs and workstations fueled the proliferation of smaller GIS for municipalities and corporations as well the rapid development of software applications that did not require costly mainframe computers (Longley et al. 2001).

Individual departments, for example planning and utilities, could develop custom map-based products and applications for their specific needs, thereby increasing efficiency and cutting costs.

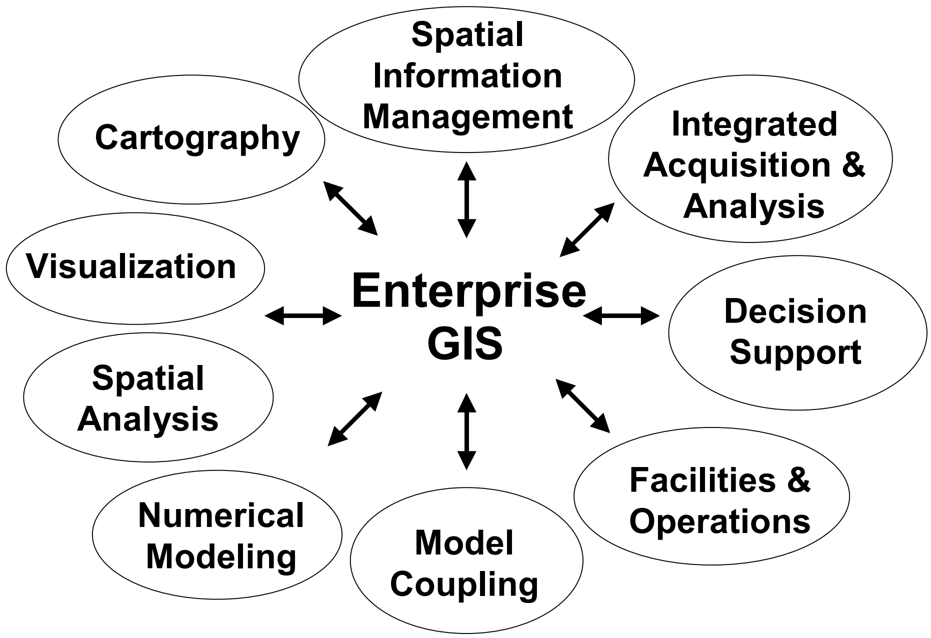

Figure 1. EGIS plays a central role of integration and access to geospatial information and analytical resources.

The recent trend toward integration of information resources through intranets and the Internet presents a challenge to the further development of local GIS teams into true enterprise resources. The use of geospatial data and analyses is growing beyond the traditional territory of individual departments, and the value of integrated data resources is increasingly being recognized at the level of corporations and national agencies (Bonham-Carter 1994). However, the project-oriented GIS model does not scale well to the needs of a large, computer-enabled institution in which GIS is used broadly for research, planning, and everyday operations (Keating et al. 2002b, 2003). For instance, when smaller project-oriented groups were combined and expected to work together on regional transportation projects, they were confronted by the challenges of redundancy of efforts and inconsistency of data quality (Peng et al. 1998). In a true EGIS, the ability to access shared GIS resources, including data, analysis software, and computing capacity, can greatly increase efficiency and flexibility to complete a wide variety of GIS tasks, and can improve the decision-making process. Enterprise design can enable a GIS user to access clean, consistent, and properly documented spatial data from institutional servers 24 hours a day without the assistance of data specialists. Such resource sharing, with concurrent access by multiple users, is foremost among the advantages of an EGIS.

Although there are a number of management approaches and institutional models that have been developed to describe a particular EGIS implementation from a management perspective (e.g., Oppmann 1999, Fletcher 1999, Lembo 1999), a complete conceptual model for the technical design of EGIS is lacking in the peer-reviewed literature. Case studies describing attempts at migration from project-oriented to EGIS emphasize the need for in-depth needs assessments, development of consistent standards and procedures, identification of core data sets, and training and participation of users and management (Coiner 1997, Lloyd 2000, Somers 2002). Practitioners cite problems in migrating large institutional (governmental, corporate) GIS to the enterprise model, including redundancies in departmental hardware, software, and databases ("stovepiping"); lack of standards; low levels of data and other resource sharing ("poor institutional behavior"); limited participation; and financial limitations for individual GIS groups (Coiner 1997, Fletcher 1999, Lloyd 2000, Somers 2002, Keating et al. 2002b, 2003). In addition, reports of successful implementations may not be in keeping with unofficial evaluations of the performance or appropriateness of the new system (Caron and Bedard 2002).

The emergence of EGIS concepts at Los Alamos National Laboratory (LANL) resulted from the May 2000 Cerro Grande wildfire that burned over 43,000 acres near Los Alamos, including over 8,000 acres of LANL land. The Cerro Grande Rehabilitation Project (CGRP) was launched at LANL shortly after the fire. The goals of the CGRP included restoring LANL infrastructure, informing the public, assisting scientists and emergency managers in assessing long-term environmental impacts, and mitigating future hazards associated with the aftermath of the wildfire.

As a result of the fire fighting and subsequent rehabilitation efforts, massive amounts of geospatial data were generated. The multi-agency Burned Area Emergency Rehabilitation (BAER) Team focused an evaluation of burn severity and potential rehabilitation on the main area of the burn in the portions of the Santa Fe National Forest and County of Los Alamos that adjoin LANL. Operational specialists from LANL concentrated on similar evaluations of Laboratory lands. Geospatial data were generated to support decisions about post-fire recovery, to mitigate floods and other hazards, and to document potential environmental impacts. A GIS component of the CGRP was charged with capturing and managing geospatial data, providing rapid access to and visualization of the data, and integrating the data into predictive models and risk assessment systems. GISLab was established to lead this CGRP-GIS effort, building on a cartographic facility that served the Environmental Restoration (ER) Program at LANL. It was quickly realized that these large projects transcended institutional divides and that a successful GIS would have to be implemented at an enterprise level.

The goal of this paper is to provide a comprehensive synthesis of the current knowledge concerning EGIS design. This knowledge includes both the conceptual foundation for EGIS design and application of these technical design principles to actual EGIS implementation. First we develop a conceptual framework for EGIS. Next we describe a proven five-step process for planning and implementing EGIS. Then we describe an implementation of the five-step framework in an EGIS for a post-wildfire rehabilitation project at Los Alamos National Laboratory, highlighting both challenges and solutions. Finally, we evaluate the success of our implementation, distill lessons and principles of broad applicability, and chart future directions.

2 Conceptual Framework for Enterprise GIS

A conceptual framework for EGIS must involve sound technical and business practices, take full advantage of rapidly evolving technology, and meet changing institutional needs. Three sets of concepts are key: 1) the role of stakeholders, 2) a complete geospatial data cycle, and 3) the design of data warehouses. In addition, a sound methodology is needed for implementation of EGIS (Witkowski et al. 2002).

GIS stakeholder roles: Complete enterprise design considers the diverse and integral roles that participants, or stakeholders, play in the process (Dueker and Butler 2000). Data providers need consistent standards and effective tools to prepare, organize, and document their data, as well as to ensure that the data will be responsibly managed. Data managers need consistent workflow procedures that ensure efficient, standardized means to manage and deliver data. GIS users need consistent mechanisms to locate and access well-documented and reliable data. Customers benefit from timely and reliable service, based on sound technical and business design.

Successful EGIS design depends on facilitating each of these participants in the context of the geospatial data cycle.

The geospatial data cycle: A complete, or unbroken, geospatial data cycle involves flows of data from data source to database, from database to applications, and, if modifications have been made, from applications back to the database, with necessary steps to ensure that data are complete, secure, documented, and accessible (Figure 2). These steps include a suite of necessary data operations: formatting, quality assurance, documentation (metadata), cataloguing, tracking, backup, delivery, and updating. Peer review of data can ensure quality and broad utility.

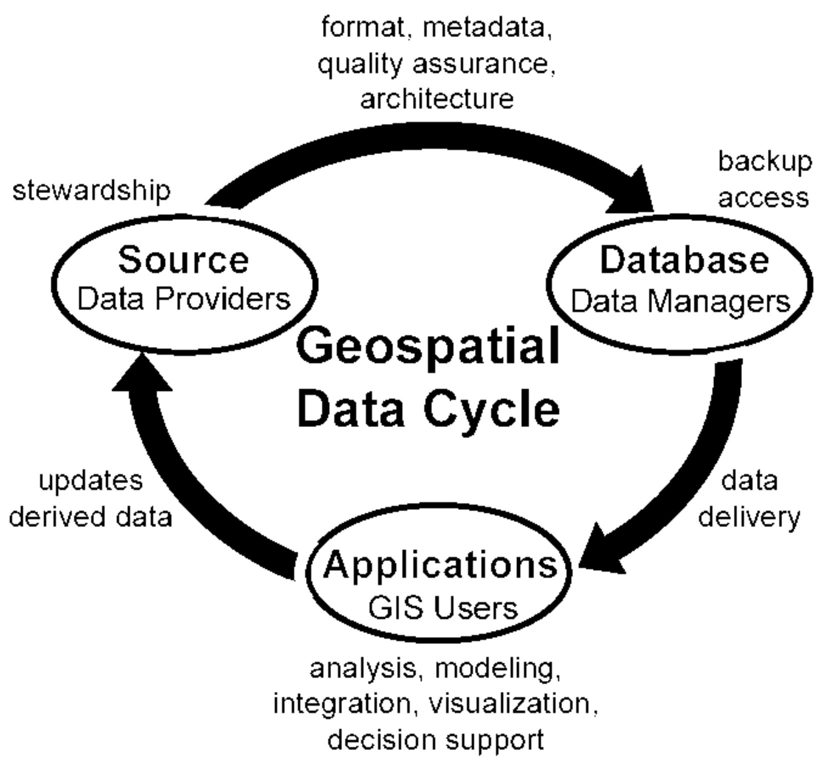

Figure 2. A complete geospatial data cycle ensures that data are complete, secure, documented, and accessible to GIS users.

Spatial data warehouse concepts: The role of a spatial data warehouse within an EGIS is to provide a clean institutional structure for the management of geospatial data and consists of three main steps: staging, storage, and delivery (Figure 3). The three steps that define data flow within the context of the work performed are as follows:

- Staging involves receipt of data and preparation for placement in the data warehouse;
- Storage involves the actual housing of data in the data warehouse; and
- Delivery involves distribution of warehoused data to GIS users.

A spatial data warehouse can provide access to shared data necessary to perform the day-to-day operations and research and development (R&D) activities.

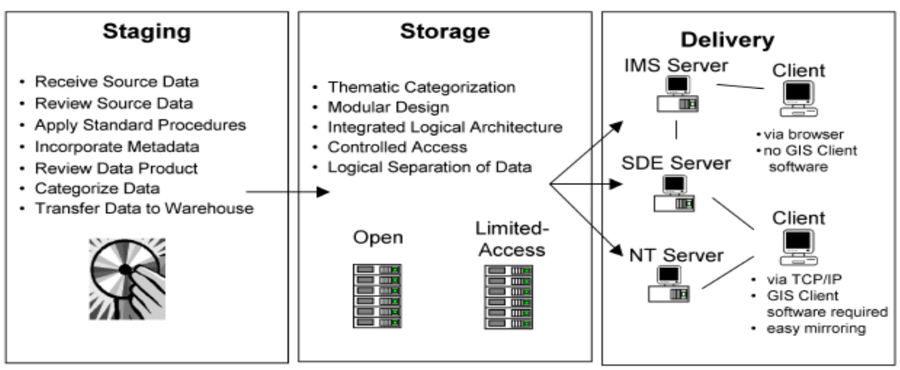

Figure 3. The spatial data warehouse includes three steps for data management: staging, storage, and delivery.

A five-step methodology of EGIS implementation: We apply a five-step methodology to implement EGIS (Dangermond personal communication 2002, Keating et al. 2002b, 2003):

1) Design specification, involving formulation of an institutional vision and goals, along with comprehensive assessment of needs;

2) Resource evaluation, including comprehensive inventory of internal and external personnel and computational resources, formulation of a list of framework or core data, and assessment of resource gaps;

3) Logical system design, including data and work flow, within the context of data staging, storage, and delivery, and including necessary policies, standards, and procedures;

4) Physical system design, including integration of servers, networks, software, and tools; and

5) Implementation, involving a plan that specifies tasks, schedule, funding, and institutional responsibilities.

This comprehensive methodology is based on proven business practices. Variations of the five-step approach have been reported in the literature (e.g., Oppmann 1999, Somers 2002).

3 Methods: The Five-Step Process for Enterprise GIS

The five-step process provides a comprehensive methodology for planning and implementing EGIS. The first two steps involve formulating a vision and strategic goals, and assessment of institutional needs and resources. The next two steps involve logical and physical design. The final step involves development of a specific plan for implementation.

3.1 Step One: Vision Formulation

A clear vision for EGIS requires an insightful vision statement and formulation of specific goals that serve the institution.

Vision statement: The first step in developing a vision for EGIS involves formulating a clear vision statement, an abstracted idealization that represents the shared values that an institution strives to achieve (Kirkpatrick et al. 2002). The vision statement keeps the process on course despite the small-scale evolution in the implementation details and helps to avoid a "reactionary" response when crisis occurs. We propose the following generic vision statement: EGIS serves the needs and reflects the values of our institution, through effective design, successful implementation, and ongoing research and development. Effective design involves careful planning. The successful implementation is the realization of this planning along with R&D to meet the changing needs of an institution. The overall goal is to serve the institution.

Needs assessment and strategic goals: At the core of developing a vision, specific institutional needs must be assessed, and a set of clear strategic goals must be formulated to address these needs. The strategic goals typically include the following: 1) achieve a complete geospatial data cycle with integrated work and data flow processes; 2) accommodate the diverse needs of an institution, including those of data providers, data managers, GIS users, and customers; 3) maximize system accessibility; 4) share key spatial data resources that are well maintained; 5) provide access to GIS technology that is intuitive, efficient, and easy to use; and 6) improve institutional efficiency while realizing cost savings. An important aspect of the needs assessment is the identification of critical processes and business operations which, if interrupted, would adversely affect the performance of an EGIS. For example, sound business operation requires periodic hardware, software, and network maintenance and upgrades, thereby minimizing downtime and maximizing the availability of an EGIS.

3.2 Step Two: Resource Evaluation

In order to successfully achieve strategic goals, a thorough evaluation of the personnel, computational, and data resources should be conducted.

Personnel and computational resources: Typically, personnel and computational resources are distributed across several departments or divisions within an institution. These people, hardware, and software may have been assembled in the past to support specialized project roles, but their contribution to a larger EGIS must be re-examined in light of the required roles (Table 1) and realistic estimates of time required. Typically, individual GIS professionals serve several roles, especially in small institutions. Similarly, computer, network, and software licensing needs must be evaluated based on realistic estimates of system usage. Hardware, software, databases, and networks should be evaluated in terms of compatibility and ability to handle the large volume of information flow required for EGIS. It is important to identify the dependencies on data and resources supplied from sources external to the institution.

Table 1. Example roles for EGIS team personnel

Leadership	Web Administration
Program Development	Customer Support
R&D Collaboration	Legacy Data Management
Project Management	Core GIS Support
GIS Administration	Project Lead
Software Development	Internet GIS
Database Logical Architecture	Database Administration
Spatial Modeling & Visualization	Training & Technical Outreach
Industry Partnerships & Technology Transfer	Decision Support
Various R&D Focus Areas	

Data resources: Three questions must be addressed concerning data resources. First, are geospatial data complete, of sufficient quality, and documented with metadata? Second, are data properly managed and secure (through tracking, backup, change control, and archiving) as well as protected from unauthorized access or modification? Third, is there an efficient process for updates and acquisition of new data? An important aspect of the data resource evaluation is the identification and designation of framework, or core data. We define framework data as the data that are necessary to support both the business and technical operations of an EGIS (Figure 1). These data may be treated differently from data of less universal importance. For example, they may be stored in an institutional data warehouse, located on an institution's priority list, and have special open access designation. Like framework data, it is equally important to identify those external data that are critical to the success of day-to-day EGIS operations, e.g., data supplied by specific government agencies. External data sources should be evaluated with respect to quality, use constraints, and availability of mechanisms for update.

Identification of resource gaps: The process of resource evaluation leads to the identification of gaps in data, personnel, and computational resources. Missing geospatial data can be compiled in an institutional priority list. Similarly, gaps in expertise must be identified and plans made to augment existing staff. Areas that may need augmentation include Internet GIS expertise, metadata development, network and server design, and large database design. Hardware and software typically need augmentation as well, since project-level systems may not be capable of handling enterprise needs for performance and availability.

3.3 Step Three: Logical System Design

An EGIS is enhanced with a well-formulated logical system design. A logical system design includes a set of work and data flow processes, defined by policies, procedures, standards, and metadata, with logically ordered and usable geospatial data. Data warehouse theory provides a sound logical system design framework consisting of three steps: staging, storage, and delivery.

Staging: Staging refers to processing and quality assurance performed when data are acquired, prior to storage in the data warehouse (Figure 3). Well-designed staging is efficient and intuitive. Within the data warehouse, designation of a staging workspace, i.e., disk space independent of workspace where data is used in day-to-day operations, enables data managers to prepare, document, and logically organize data. A staging workspace ensures that new data are processed and properly incorporated following standards and procedures as part of the geospatial data cycle, thereby maximizing the quality and completeness of the data to be stored.

Storage: Storage refers to the configuration and logical institution of databases and directories where data are stored. Spatial data can be stored either in a series of related or spatially indexed tables within a database management system (DBMS) or as flat files (individual files within a hierarchical directory structure). Depending on GIS requirements and use, both flat file and DBMS storage methods can provide efficient means to store and access spatial data (e.g., Bernhardsen 1999, Xiong and Lin 2000). GIS Data are often organized by spatial domain (e.g., region, site, sub site, etc.), project, or theme (e.g., hypsography, roads, hydrology, etc.) (Bernhardsen 1999, Witkowski 2002).

Delivery: Another critical component of the spatial data warehouse is the efficient and reliable delivery of spatial data via a computer network (cable or wireless) from storage to multiple customers. Delivery is accomplished by mapped networked drives containing flat files, by websites containing data, or by a DBMS providing quick and efficient access to data through a series of queries. Data delivery can be made efficient by employing a multi-tier architecture in which separate servers are used for file serving, applications, and processing (Figure 4).

Efficient data sharing at an enterprise level requires uniform policies, standards, and procedures for data quality, metadata, and access (Keating et al. 2002b, 2003). In contrast to ad hoc personal exchanges, it has been shown that formalized processes and procedures for data exchange produce a better flow of information (Pinto and Onsrud 1995), as a result of a better understanding of individual responsibilities and expectations. Policies govern, in general terms, how data will be documented, managed, and made accessible. Standards provide specific requirements concerning format, structure, and content; these requirements help ensure that the integrated data management system does not break down and revert to multiple independent, stovepipe GIS within the institution (Somers 2002). Procedures define work processes and detail the steps applied to meet the requirements of policies and standards, for example to assign names for data layers, to ensure complete and consistent metadata, and to provide for access control when required by data providers.

Metadata are the data that describe the content, quality, condition, and other characteristics of spatial data. Metadata enable EGIS stakeholders to determine data availability, fitness for use, access, and source (FGDC 1998, OMB 2002). Goodchild (1995) argues that the impact of metadata on the use and value of spatial data needs further evaluation. However, it is clear that access to metadata enhances data sharing across institutional boundaries, which improves institutional effectiveness and decision-making ability (Pinto and Onsrud 1995).

3.4 Step Four: Physical System Design

The physical system design provides the framework and configuration for the integration of core hardware, software, and computing infrastructure necessary for efficient and reliable data sharing. Unlike the logical system design, which is purely conceptual, the physical system is highly dependent on the integration of specific software and hardware components (Bernhardsen 2000). Brady (2000) advocates a three-tier system for EGIS: 1) storage / database, 2) application servers (spatial database engine, mapserver, tools, web), and 3) clients (ArcGIS, browser, generic workstation) (Figure 4). A primary goal for physical system design is to achieve high availability of networks, servers, and storage devices through redundancy, clustering, and fail-over algorithms (Brady 2000). High availability for servers has been quantified as roughly eight hours (or less) of downtime per year, but this value depends on the combined performance of all system components (Brady 2000). Each institution must develop its own criteria for high availability and consider all potential causes of failure, impact to data access, and a response plan to minimize system downtime.

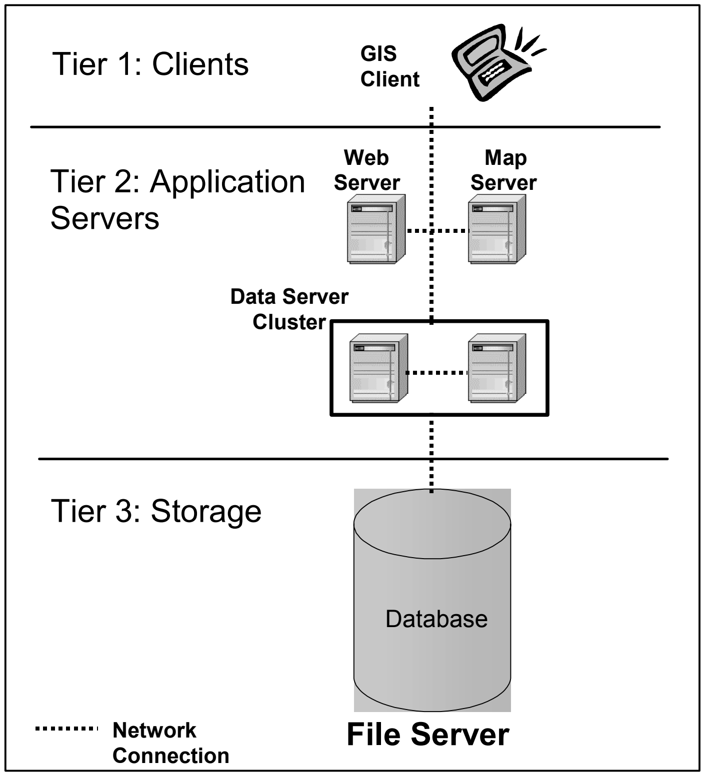

Figure 4: The three-tier architecture provides high availability of GIS software and data, and each tier can be expanded and fine-tuned to allow the EGIS to evolve and scale.

Networks: Fast and reliable networks serve as the backbone for EGIS, both local area network (LAN) and wide area network (WAN). Network failures and disruptions in service can cripple an EGIS. Some critical systems rely on redundant networks to ensure high network availability (e.g., Choi and Somani 1996). Redundant networks consist of either a pair of separate network cables that connect to a client machine (housing two separate network cards) or one network cable and a means for a client machine to communicate (wireless) with the redundant network. Redundant networks can be expensive, especially if there are multiple levels of security and different data transmitted on each network.

Servers: Many GIS applications require computationally fast and reliable servers that are able to process numerous concurrent data requests from users. Whether applications are "thin" (execute on server) or "fat" (execute on client machine), high availability is essential. High server availability can be accomplished through the use of a multi-tier physical architecture and server clustering, whereby users access a virtual IP address for the cluster and are directed to the server in the cluster that is running the fewest processes (Brady 2000, Gan et al. 2000, Cardellini et al. 2001). This method helps to ensure a balanced load of processes on the servers and minimizes software configuration on client machines. Server clustering also minimizes the impact of server failure by monitoring server performance, detecting failure, and redirecting processes on the failed server to other functional servers without an interruption of service to clients (Brady 2000, Gan et al. 2000, Cardellini et al. 2001). Availability can be further enhanced by the use of a "hot standby", a server that normally runs no processes but which can take over processing in the event of server failure (Brady 2000). A major advantage of server clustering is scalability: more servers can be added to a cluster as the demand increases, without a significant reengineering of the physical system design.

Storage devices: In addition to servers, efficient redundant storage devices, such as a Storage Area Network (SAN), are necessary to store the vast quantity of data used in an EGIS. Highly reliable, redundant storage devices usually consist of RAID (Redundant Array of Independent Disks) systems or similar technologies that provide a fail-safe mechanism in the event of disk failure. In order to address diverse user needs, it is important that access to the storage device be flexible enough to communicate seamlessly with multiple operating systems, such as UNIX and

Microsoft Windows. These systems typically include server software to provide frequent local data backup. An additional backup or mirror of data at an off-site location is another level of redundancy that may be a lifesaver in the event of disaster (Mynard et al. 2003).

EGIS software and custom software tools: An invaluable component to the design of the physical system is system software used to monitor network, server, and storage device performance; detect hardware failures; and provide automated methods of fail-safe protection. EGIS software enables users to access data and analysis resources. Stakeholders may choose to use different brands of GIS software depending upon their needs and preferences. However, uniformity in the GIS software used can save institutions money (e.g., through site license agreements) and ensure compatibility (e.g., compatible file formats).

Custom software tools and applications must be developed when commercial software fails to meet the needs of stakeholders. Commercial software typically does not provide functionality for such activities as work and data tracking and web-based management of metadata. Many of the problems encountered with custom software development, such as code design, different development environments, testing, and versioning can be addressed with effective software development and management practices (Brown 1996).

An important aspect in developing custom software tools is communication with the larger software industry. Technology transfer provides a feedback mechanism from developers of custom tools to commercial GIS software companies.

Software and hardware testing environment: Since EGIS technologies are rapidly evolving, it is important to test new hardware and software. A "testbed", or testing environment, should be created independent of critical business systems so that software development does not adversely affect daily GIS operations. The testbed should accommodate a thorough evaluation of all computer systems and software before they are placed online.

3.5 Step Five: Implementation

The actual implementation of an EGIS translates a technical design into a working system via a series of discrete tasks, according to a schedule, and depending on adequate funding and sound business practices.

Tasks: The specific tasks involved in implementing an EGIS include aspects of both the technical design and the business models are as follows:

- Evaluate and acquire computing infrastructure and budget;
- Develop policies, procedures, and standards that constitute a spatial data management plan;
- Determine personnel needs (existing expertise, hiring);
- Install and configure system components; and
- Develop customized applications.

For many institutions the design and implementation of an EGIS can best be accomplished by hiring an outside consultant to supply the needed expertise.

Schedule: While it may be desirable to transition to EGIS quickly, a phased approach is often more practical. Such a phased implementation is dependent on budgetary restrictions, institutional needs and priorities, project and workflow demands, and other business and technical realities. Part of the implementation plan involves the development of specific timetables. It is important to be adaptable during the implementation process, and to accommodate technology changes, as well as the changing needs of the institution.

Funding: In large institutions GIS is typically not directly funded, but rather it develops as a by-product of a particular project or initiative; for example, in the U.S. Department of Energy (DOE) complex, advanced GIS capabilities have grown out of environmental restoration programs. As a result of this initial orientation to serve project needs, these focused GIS may not be sufficiently responsive (or fiscally stable) to meet other needs of the larger institution. In contrast, stable funding for EGIS can be ensured through four mechanisms as follows:

1) Direct allocation provides funds to build and operate an enterprise facility as a budget line item;
2) Fee-for-service places GIS as a viable business within the institution, wherein expenses are balanced by billing customers for work performed;
3) Overhead funding generates funds as a set proportion of external funds that come into an institution; and
4) Facility fees generate funds not as standard overhead but rather as specific fees paid by projects that use or potentially could use the facility.

Institutional responsibilities: The institutional structure into which the EGIS is nested can pose one of the most difficult challenges to implementation (Keating et al. 2002b, 2003). First, institutional responsibilities must be clearly defined, either by strong consensus among the teams involved or by management fiat. Second, it is essential during the transition to EGIS that the technical capabilities be understood and served by management. The challenge is to avoid a disconnect in which management does not have enough information about what GIS is or why it is important to the institution. The institution must build a "team of two" of technical expertise and management support (Dangermond personal communication 2002, Keating et al. 2002b, 2003) to move EGIS forward. Third, there must be an institutional champion for GIS (Oppmann 1999), for example a Geographic Information Officer (GIO). This office or individual ensures a common vision during and after the transition and provides leadership and authority.

Research and development: Ongoing GIS research and development (R&D) is essential to meet changing needs of day-to-day operations and to incorporate changing technology.

Communication, outreach, and training: As the technical components of an EGIS are assembled, the human components of the system must be involved as well. Reaching beyond the GIS and IM domain experts, this effort involves teaching people in the institution what GIS is and how to utilize its new power in the institution (e.g., Coiner 1997, Oppmann 1999, Somers 2002).

External review: In addition to internal stakeholder input, establishment of an external review or advisory board provides a mechanism for obtaining technical and institutional guidance from GIS experts and leaders.

Measures of success: The success of an EGIS depends on achieving essential goals while maintaining institutional values. Each institution must develop a set of metrics based on its specific goals; in essence a measure of how well the vision becomes reality. Success is evaluated relative to elements and criteria that measure system performance.

4 Results: A Prototype for Enterprise GIS

4.1 Vision for GISLab

GISLab was established to provide EGIS solutions to accommodate the diverse needs of the CGRP and other projects at LANL by acquiring, storing, and providing access to high-quality, well-documented data for post-fire rehabilitation, hazard mitigation, and environmental monitoring efforts.

The diversity of project and operational data needs, as well as points of view regarding spatial data exchange and storage, posed a serious challenge to the CGRP-GIS effort. The needs of the stakeholders were assessed collectively through stakeholder meetings and individually through a web-based consensus-building and conflict-clarification tool. This consensus tool provided a means to assess the positions and attitudes of the various CGRP stakeholders and to help shape the design of the GIS to meet their needs (Keating et al. 2002a). The online, open-response survey used in the tool harnessed the collective intelligence of fire rehabilitation stakeholders to identify areas of conflict and agreement and to provide documentation for GIS design. The results highlighted a clear division of labor between operations personnel, who were involved with infrastructure rebuilding and public information, and environmental monitoring and research workers, who were concerned with evaluating environmental impacts, numerical modeling, and flood control. The main GIS needs identified by the stakeholders included topographic data, a central data repository, remote sensing data, predictive model results, and data on post-fire environmental changes. The perceived need for a central data repository was most strongly correlated with the need for infrastructure and topographic data and predictive model results. Concerns were voiced about potential problems with data access, ownership, and maintenance; costs; and redundancy.

The goals of GISLab were established to meet several institutional needs. First, a **public website** was required to publish the varied geospatial data and cartographic products of the BAER Team: the progression and severity of the fire, forest mortality, and proposed treatments, such as slope stabilization and seeding (http://www.cgrp-gis.lanl.gov). Second, a multi-use **spatial data warehouse** was required for storing geospatial data and associated tabular data related to the fire, rehabilitation, environmental monitoring, and restoration efforts. These data included terrain, hydrology, infrastructure, vegetation, geology, ecology, archaeological sites, and areas of potential environmental impact, as well as remotely sensed images. The spatial data warehouse had to provide efficient storage and delivery of the data and to incorporate standards for quality assurance, documentation (metadata), and security. The data would be accessed in several ways, including direct GIS client software connections to the database, network transfers of flat files, and via the Internet through an Internet Map Service (IMS). Third, a **web-based metadata**

catalogue was needed to provide a tool for researching available data sets. Fourth, the GISLab team would provide **expert spatial analysis and cartography services** to the CGRP and other projects.

4.2 Resource Evaluation

GISLab was built on a foundation of existing GIS resources, including the Facility for Information Management, Analysis, and Display (FIMAD). Other LANL GIS teams (e.g., facilities and infrastructure, planning, environmental monitoring and restoration, and earth science research) were included in the resource evaluation as sources for fire rehabilitation data and as users of the EGIS.

FIMAD databases contained LANL environmental data, including terrain, infrastructure, land ownership, vegetation, hydrology, geology, ecology, and archaeology. These data were augmented with data specific to the Cerro Grande fire, including BAER Team fire data layers and maps that depicted fire-fighting activities, burn severity, vegetation mortality, hazards, and proposed treatments. In addition, up-to-date LANL infrastructure data were acquired from Johnson Controls Northern New Mexico (JCNNM), including roads, utilities, buildings, etc. Finally, data were acquired during the post-fire rehabilitation and hazard mitigation work, including maps of forest thinning activities, potential floodplains, and erosion potential.

As the EGIS tools were being developed for the cross-institutional CGRP-GIS, efforts were made toward the establishment of institutional data clearinghouse for LANL, that is, a larger institutional GIS framework for basic operations and research in the entire institution. As part of that work, a newly formed LANL GIS Technical Steering Committee began to develop a list of core institutional data, including terrain, hydrology, roads, buildings, utilities, vegetation, threatened and endangered species, and geology. These data may reside on a central server or on distributed servers, but a central metadata clearinghouse; standards for data quality, format, and documentation; and access protocols will increase the efficiency of GIS work at the institution.

The number of GIS personnel in FIMAD were not sufficient for the implementation of the GISLab EGIS. As a result, four additional staff members were hired. Computational resources were augmented as well, including data storage devices, web and application servers, and desktop workstations.

The rehabilitation and hazard mitigation activities required a level of geospatial data detail that was often not available in pre-fire datasets. For instance, high-resolution digital elevation models (DEMs) were required for modeling flood and sedimentation potential and for identifying areas for prescribed forest thinning. As a result, precise LIDAR elevation surveys were conducted to construct one- and four-foot DEMs. Revised land cover classifications were required both for studies of fire effects and for planning rehabilitation efforts. Finally, data describing and quantifying post-fire treatments performed and hazard model results were needed in the spatial data warehouse.

4.3 Logical System Design

The GISLab design for EGIS, like various others (Lembo 1999, Oppmann 1999, von Meyer 1999), is focused on the flow of information and data to the GIS user from centralized storage

(Figure 3). We emphasize the integration of workflow and dataflow processes and stakeholder roles in the concept of data staging, storage, and delivery.

In staging, data are placed in a workspace organized according to consistent data architecture, based on project, spatial domain, and theme (Figure 5). Here data are processed to ensure standard formatting, quality assurance, and adherence to metadata standards. As a new or revised dataset is accepted, its lineage is documented in a new or updated metadata file and it is assigned a unique tracking number. A copy of the data is stored in a source archive, organized hierarchically in directories by project and request number, a unique identifier. Another copy is placed into a staging directory for processing, where it is sorted by project, spatial domain, and theme. A quality assurance check is performed to ensure that the data are projected using standard model and datum, and FGDC-compliant metadata are developed.

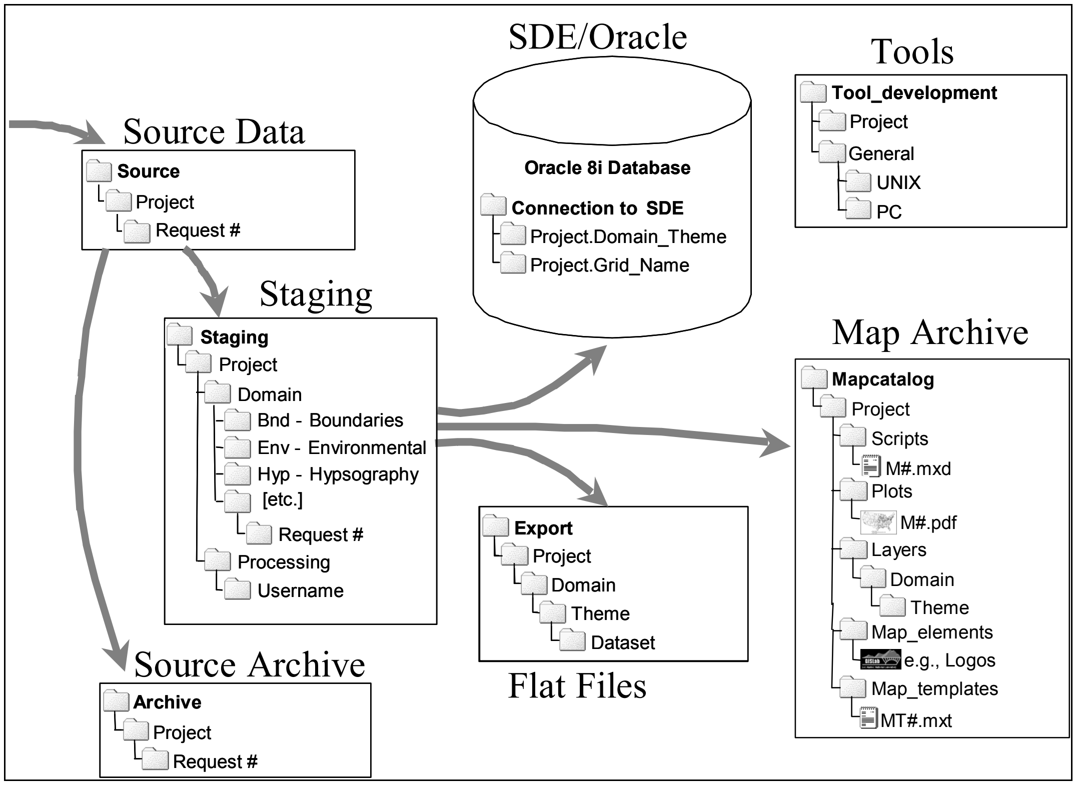

Figure 5. The logical data architecture enhances data flow from source to staging, staging to storage, and storage to data delivery via flat files and relational database access. Archives of source data and products (maps, plots, templates, etc.) are stored for future reference. Tools (scripts, source code, and executables) are also stored in a standardized format.

In order to address problems of compatibility of various spatial datasets of different scales, the logical data architecture is designed to emphasize the thematic categorization of spatial data into scale-dependent domains. For example, a spatial dataset of LANL surface geology is sorted into the appropriate local spatial domain LSR (meaning "laboratory and surrounding region") and then is stored in the GEO (geology) thematic category. The data structure expands hierarchically to handle data covering larger aerial extents, for example regional watersheds. In this manner spatial data can be categorized quickly and can be easily searched in a logical fashion. The structure is efficient and intuitive, reducing the time spent by analysts locating datasets and ensuring consistent coordinate systems.

The logical data architecture of our EGIS not only specifies categorically how data are catalogued and organized, but it also specifies how data flows in and out of our spatial data warehouse. This data and workflow specification is accomplished by a robust and efficient tracking system. We designed and implemented a web-based tracking system, called the Request System, which provides a user interface and management tools for a series of Oracle (http://www.oracle.com/) tables that contain information concerning the data and work flow (Linger et al. 2002a). The Request System is accessed through a JAVA-driven web interface through which users can enter, submit, and query previous requests.

Conceptually, the Request System is part of GISLab workflow processes; however, it is also integrated in a number of ways with the data flow and logical data architecture. A unique identifier (request number) identifies requestor contact information, charge code, due date, type of work requested (cartography, spatial analysis, data processing, etc.), and special instructions. The same request number is used in directory and file names for the raw and processed data as well as any maps produced from them (Figure 5). When there is a request or a need to update or modify an existing dataset, the request is assigned a new unique identifier. All versions are archived according to a file architecture based on these identifier numbers. The status of the work, the location of the data and related maps, and the lineage of the original data are thereby intimately related. To date, over 11,000 original maps have been produced, recorded, and reproduced using this application.

The Request System is integrated with map gallery tables, also stored in Oracle, which are categorized by the domain and map extent defined for the various maps published. In this way, work requests are recorded, progress is tracked, and the product is archived in a systematic way. For example, once a cartographic request is completed, the map extent information is recorded, and the map is archived. A user can later use a query tool on the map gallery web page to find products by aerial extent or request number, and maps in Adobe Acrobat pdf format are displayed and made available for direct download. The Request System primary table is also available from ESRI's spatial database engine (ArcSDE) and can be used to join attributes from various spatial layers that are assigned a request number. This process enables users to track changes made to particular spatial features within a spatial data layer, providing a mechanism for feature level metadata and lineage tracking.

Geospatial data are stored in the actual data warehouse according to the same architecture used in staging. GISLab employs redundant file storage formats, both to protect data from loss or corruption and to honor specific access requirements of data stewards and users. The primary storage format is within Oracle 8i database tables, accessed through ArcSDE. Current off-the-shelf software (Oracle, ArcSDE) supports storage of vector and raster spatial data as well as text and tabular information. While the current version of ArcSDE (8.3) does not allow true clustering of data servers, instances of the databases and SDE on separate GISLab servers (Figure 6) provide redundancy (mirroring or backup) and access control to sensitive and project-restricted datasets. This measure of control is further developed through database user "roles" (a type of user profile) that limit access to groups of datasets. Geospatial data are also stored in "flat files" within hierarchical, shared "export" directories (Figure 5). Data security, in terms of access, is ensured by permissions settings. Data backup or mirroring of these directories is straightforward through typical network administration tools. Efforts are currently underway to develop a remote backup data storage location, possibly elsewhere within the DOE complex.

Data are delivered from servers by means of shared directories or relational database connections. The ArcSDE connection to ESRI's GIS client software (ArcGIS, ArcIMS) provides enhanced data structures and access efficiency, and the ArcGIS client software is designed to view metadata and tabular data that are associated with the spatial data. Users can also find and copy geospatial data files across the Unix or NT networks. This method provides lowest-common-denominator access to basic files in standard format, but it does not ensure that links among related files (and to metadata) are maintained.

In addition to the network access methods, subsets of CGRP geospatial data are available from websites through web browsers. Compressed files of public BAER Team data, maps, and metadata are available by direct download from data catalogue and map gallery pages on the CGRP-GIS website (http://www.cgrp-gis.lanl.gov). This website also provides a gateway to interactive map services built with an Internet map server (ArcIMS). Basic map services allow visualization of pre-packaged sets of geospatial data layers (vector and raster) and metadata. Users can zoom and pan maps, turn on and off layers, and query the attribute tables associated with the geospatial data. More complex IMS implementations provide feature-streaming capabilities, in which data are streamed to the client machine to allow advanced GIS interactions like editing, annotation, and data exporting. In short, the Internet extends the data distribution capabilities of the EGIS by providing a platform for a metadata clearinghouse, direct download of public-access data and maps, and interactive visualization of related spatial data.

The hardware, software, geospatial data, and people involved in the EGIS are bound together by rules and agreements for data access, standards for quality, and accepted procedures for work and data flow. While avoiding draconian regulation of GIS activities, we developed policies, standards, and procedures that establish basic methodologies for work and consistency in the products. At the same time, these agreements lay out responsibilities and expectations for the varied roles of people involved with the GIS and tie together work and dataflow.

Most notable among GISLab policies is the Data Access Policy, which specifies provisions for both data providers and data users. The policy describes how the data in the warehouse will be managed, including quality assurance checks, the requirement that FGDC-compliant metadata accompany all geospatial data in the warehouse, and an offer for assistance in preparing data and metadata for submission. Data providers may specify restrictions on data access and use. Users agree to abide by restrictions on access and use, to properly cite data sources, and to provide updates or corrections to the data managers. All users must sign a data agreement. The policy also assures users that the spatial warehouse will be properly maintained such that 1) spatial data and metadata are available to authenticated users via the internal LANL network, 2) all data are backed up on a regular basis, and 3) limited-access data are available only to those data users approved by the appropriate data provider. Metadata for both open and limited-access data are openly available.

The GISLab Metadata Standard follows the FGDC metadata standard (FGDC 1998) based on data documentation requirements of OMB Circular A-16 and Executive Order 12906 (Clinton

1994, OMB 2002). Beyond compliance with FGDC standards, the GISLab metadata standard requires LANL-specific data fields (e.g., classification, etc.). We provide web access to a GISLab metadata template as well as to FGDC resources and examples.

Procedures outline the accepted work practices and methodologies to implement the policies and standards. Specific GISLab procedures include Change Control, Data Processing, Workflow, Cartography, Metadata Preparation, and Website Design and Maintenance (http://gislab.lanl.gov/policies.html and http://gislab.lanl.gov/guides.html).

One of the first procedures applied to the dataset while in a staging location is a change control procedure. The change control procedure addresses data stewardship concerns — who is authorized to make changes, how the dataset will be versioned, and ultimately how the changes will be implemented. This procedure requires determination of who the data steward will be and what type of process will be in place to review future changes to the dataset. It is often the case that data can change very rapidly and thus make a change control procedure very cumbersome and oppressively rigid. It is recommended that only "Official" or published, versioned data be processed under a change control procedure. This recommendation can prevent the use of the latest data that has not been processed, but it does provide confidence that available data has been thoroughly reviewed.

For all data stored in the staging area, we construct FGDC-compliant metadata using ESRI's ArcGIS ArcCatalog tool. Recognizing that metadata development can consume a tremendous amount of time, we have streamlined the process by developing metadata templates. Thus, we can rapidly apply repeated information and focus on documentation unique to each spatial dataset. In addition to saving time, templates provide a consistent look and feel to metadata. In an effort not to be dependent on a single tool, we publish and serve HTML metadata files on our website as well as store XML files associated with the data and accessed using ArcSDE. This method allows users of older versions of GIS software, such as ESRI's ArcView 3.x software, to see and download the metadata for geospatial datasets of interest.

4.4 Physical System Design

The GISLab physical system design is based on the integration of several high capacity RAID-4 storage hardware devices with an Oracle 8.i Relational Database Management System (RDBMS) and the ArcSDE spatial database engine.

The backbone of the GISLab EGIS is a series of high-performance networks consisting of CATV (Hybrid Fiber Coax) cables integrated and interconnected via the TCP/IP protocol. Stakeholders with the appropriate access permissions, IP address, network card, and port connect their desktop computers to the network to use the shared data resources contained in the EGIS. In an effort to reap the benefits of pooled or shared resources, some stakeholders, data managers and users rely on their network connections to access GIS software licenses via a license server. This server allows for a number of users to share not only data, but also the other resources, essential to conduct day-to-day GIS operations.

We employ a multi-tier server configuration, consisting of two non-clustered servers (Sun Microsystem's 420R Workstations), windows-based web server (Microsoft IIS) and Internet map server (ArcIMS); and a file server to serve data to clients in the EGIS (Figure 6).

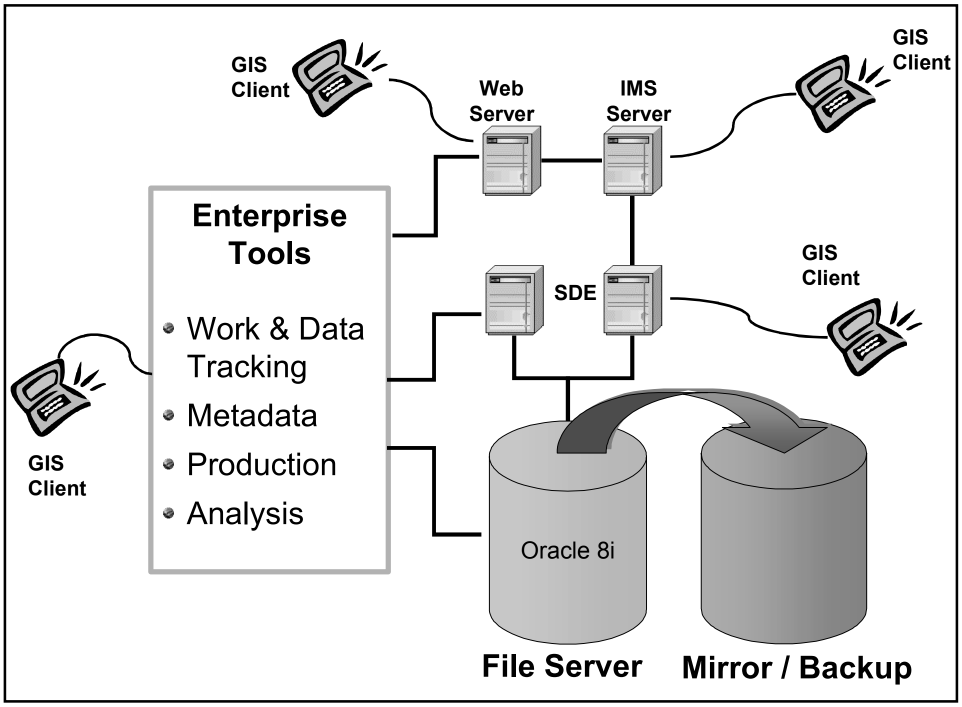

Figure 6. The GISLab multi-tier spatial data warehouse physical architecture includes servers and software for data storage and delivery and enterprise tools.

The foundation of our physical data storage architecture is a Network Appliance's NetApp F760 storage device that has the capacity to store over 3.3 terabytes of spatial data. Our storage device utilizes a RAID system and intelligent backup mechanisms and provides the ability to access data from multiple file protocols (Microsoft's Windows NT, UNIX).

In addition to the Request System, GISLab utilized commercial software products and developed several others to design and implement EGIS.

- **Map tracking:** GISLab uses the Mnum application, an interface to track cartographic products and their attributes. Mnum utilizes Microsoft's Access database to store information about each map produced, including the title, spatial extent, projection, component data layers, and most importantly, the request number that ties map data to the Request System. Future enhancements call for development of a web interface for examining map data and migrating data storage to the same Oracle database used for requests.

- **Cartography:** The Rapid Mapping Tool (RMT) was designed for rapid map production (Linger et al. 2002b). The tool uses the new ArcGIS architecture to replace an aging AML script-driven system. The RMT adds functionality 1) for easier access to many of the map-making functions provided by the core ArcGIS product, and 2) for storing key map information, such as data layers used and map extent.

- **Data modeling:** GISLab utilized a visual data-modeling tool, ERwin v4.0 (http://ca.com/), to design and generate the Oracle database used to store information about each work request. A logical data model was developed that included four subject areas: 1) cost information, such hourly rate and hours worked; 2) personnel information, including who requested the work and who completed it; 3) a data inventory; and 4) information about each map produced. The logical data model also includes the relationships and business rules among these subject areas, by containing all the primary keys, foreign keys, alternate keys, indexes, sequences, and triggers.

- **Batch processing:** GISLab developed a batch tool, Visual Basic extension to ESRI's ArcCatalog and ArcToolbox, using ESRI's ArcObjects. This batch tool provides utilities for generating, editing, and searching metadata, and to manage large raster datasets containing hundreds of individual tiles. Functionality includes selection of multiple objects in ArcCatalog, application of complete or partial metadata templates to those objects, export of metadata in many schemas, definition of projections, and loading of raster images into SDE using a standard naming convention.

- **Data backup:** In the case of accidental loss of data, the Network Appliance software provides "snapshots" or short-term incremental backups of the data on the system. Snapshots of data are taken every 60 minutes. This capability is of critical importance for rebuilding the database in the event of a software or hardware failure.

- **Consensus building:** A custom web-based tool was developed for consensus building and conflict clarification in large stakeholder groups (Keating et al. 2002a). This tool was used during the design and implementation of the CGRP-GIS by collecting, synthesizing, and documenting stakeholder views and concerns regarding data and GIS service needs, data warehousing, and potential costs.

GISLab employs a suite of testbeds to ensure that software and hardware function properly. A Sun 420R Workstation is used as a testing environment for SDE and Oracle; numerous Access databases are used to test database design concepts before use; several PC workstations are assigned to test beta versions of GIS software; Microsoft's Visual Source Safe is used as the software configuration management tool; and Sun Microsystem's Forte is used for Java code development.

4.5 Implementation

The planning stages of the CGRP EGIS project included definition of major tasks, development of the schedule for the two-year project, and assignment of roles for personnel (Table 1). The annual project budget provided limits on spending for each aspect of the effort (hardware; software; custom applications; policies, procedures, and standards, data acquisition and loading; etc.).

Other important components of the implementation included communication, outreach, and training to increase the awareness and use of GIS in the institution. The CGRP-GIS project was publicized in brochures, a touch-screen kiosk installed in public places, a public website, and oral presentations. GIS stakeholder communications involved face-to-face meetings, the LANL GIS User Group, and a web-based consensus-building tool. The challenge of gaining participation and buy-in can be overcome through training and familiarization aimed at all levels in an institution (Oppmann 1999, Somers 2002). Training sessions for specific EGIS software and tools were coordinated by GISLab, including web-enabled GIS (ArcIMS), metadata preparation, Global Positioning System (GPS) tools, and image processing (Imagine, Erdas). Finally, GISLab worked to raise the awareness and appreciation of the power of EGIS by sponsoring visits and public lectures by national GIS experts.

The EGIS designed by GISLab was required to ensure a complete data cycle through the design and implementation of a data management system with data access policies, consistent standards and procedures, efficient storage design, and multiple access methods. The operational success of the system can be evaluated in terms of tangible criteria such as system performance (e.g., server downtime), number of users, number of data layers loaded, maps and analyses produced, funding stability, and increases in productivity (Table 2).

Table 2. Metrics for evaluating the success of the GISLab EGIS (based on calendar year 2002).

Element	Criteria	Results
Implementation of Physical System	Hardware installed	3.3 Tb file server, web server, and map server
	Software installed	ArcSDE 8.2, ArcIMS 4, ArcGIS 8.3, Oracle 8i
	Custom software developed	Request System Rapid Mapping Tool Batch Processing Tool
System Performance	Server down time	< 1 Day/yr
	Data transfer rate	11 Mbps (Max)
Utility of Spatial Data Warehouse	Number of data layers loaded and size	1419 SDE & 2304 Folders Total > 85 Gb Served
	Completeness of data and metadata	100% Complete Data w/FGDC Compliant Metadata
	Number of users	> 50/yr
Website effectiveness	Number of website visits	> 2000/yr
Workflow standardization	Policies implemented	Data Access, Contracts
	Standards adopted	Metadata
	Procedures implemented	Change Control, Data Processing, Workflow, Cartography, Metadata Preparation, Website Design and Maintenance
Productivity of GIS Services	Number of work requests completed	639/yr
	Number of maps completed	2179/yr
Funding stability	Annual budget	~ $2 million/yr
Institutional Benefits	Enhanced awareness of GIS	5 Training Seminars, 3 invited speakers, bi-weekly tech steering committee meetings, quarterly user group meetings
	Increased internal data sharing	7 LANL divisions use system
	Enhanced data usage by external stakeholders	> 10 Data Usage Agreements Signed and in place.
	Elimination of redundancy	Not Quantified
	Increased productivity	Estimated 35% Decrease in time spent searching for data
	Cost saving to institution	Not Quantified

For example, the log of access to the data warehouse via SDE serves as a tangible means to track system usage (Figure 7).

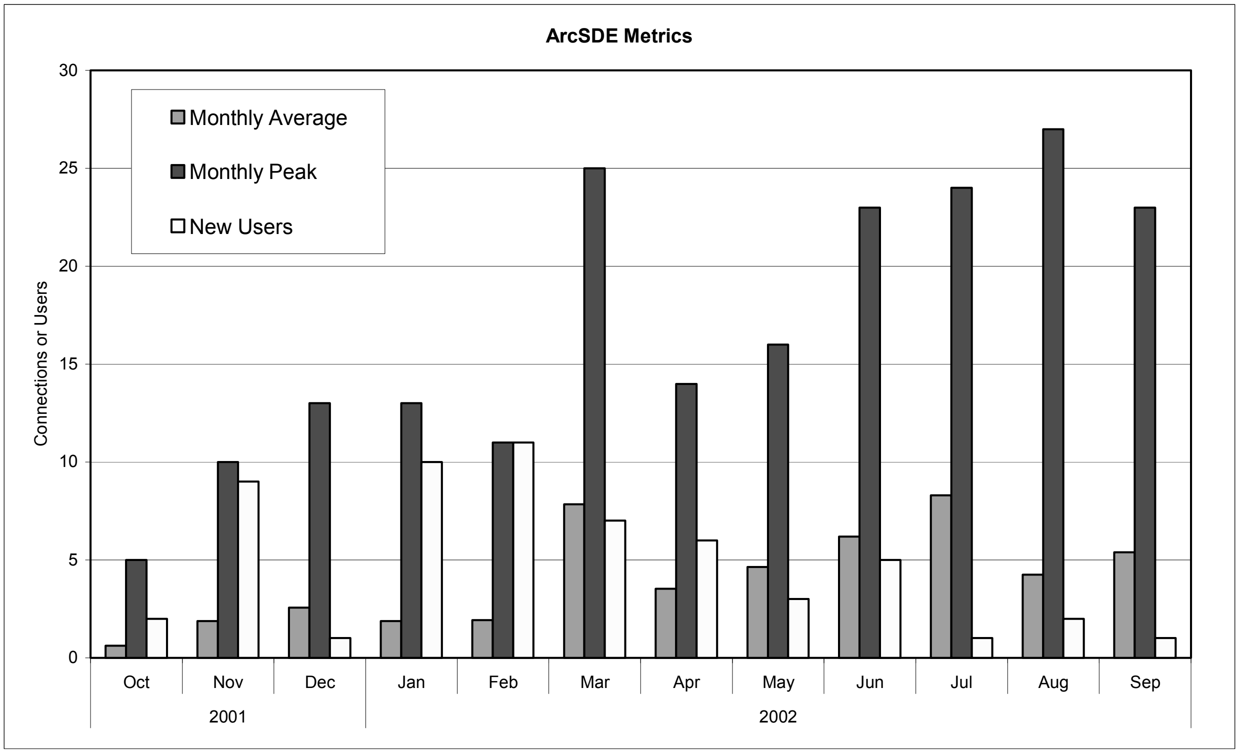

Figure 7. System usage as measured by ArcSDE access, over the course of one year.

In addition to the tangible elements listed in (Table 2), the following was observed in developing our EGIS: 1) The needs assessment and design efforts raised awareness of the importance of GIS at LANL in the areas of facilities and operations, emergency management, environmental monitoring and restoration, and earth science research. 2) At the same time, Federal standards for metadata (FGDC 1998) were being implemented across the DOE complex (e.g., Bleakly and Lee 2001, Rush 2001), and the concepts for a GISLab metadata clearinghouse for fire rehabilitation and environmental restoration data, consistent with national standards, were useful at the scale of the institution as a whole; 3) progress made by GISLab in establishing an efficient spatial information management solution focused scrutiny on the inefficiency, redundancy, and incompatibility of the independent project GIS structure within the LANL complex (Keating et al. 2002b, 2003).

5 Discussion

EGIS is constrained by the evolution of two dominant forces: technology and management. The enabling technologies -- fast computers, high-capacity storage devices, and high-speed networks -- are now relatively mature and readily available, but integration of these technological components in an effective enterprise design is still at an early stage of maturation. Conceptually, EGIS is more a management approach to resource allocation and spatial data management than the collection of technologies that enable it. EGIS requires customized design based on the goals and needs of the specific institution. The lack of a "turn key" or "one-size-fits-all" solution challenged us to develop a design that focused primarily on issues such as how to track data and work flow, rather than on more global management issues, such as how to ensure financial viability of the system. We identified three key technical elements: 1) stakeholder roles, 2) a complete geospatial data cycle, and 3) data warehouse concepts.

In the development of an EGIS, it is often easy to concentrate on issues related to the flow of data without first assessing the roles of stakeholders. Early in the process we recognized this first key element, that stakeholders, and more importantly the interoperable technical roles they play and the services they provide, would determine the success of the system. Failure to provide for critical stakeholder roles can result in an incomplete and stalled system. Good design is largely determined by the interoperable roles or services that stakeholders fill and provide.

Organizational or "Institutional" interoperability (Fletcher 1999)--for example, the ability of cartographers to produce quality maps- requires that high quality data be available from data providers. In addition, we found that the foundations of our EGIS are only as strong as the unified coordination, collaboration, and consensus achieved among the various GIS stakeholders within our institution (Keating et al. 2003). A GIO can provide important leadership and coordination for GIS stakeholders, where as an external advisory board can offer strategic and unbiased guidance. Good communication and data sharing among stakeholders are paramount to success. GIS stakeholders who have a vested interest in the success of the system become excellent ambassadors of the system.

The second key element, the concept of a complete, or unbroken, geospatial data cycle ensures that all necessary elements and processes of an EGIS are present. For example, sharing of data is difficult when metadata are lacking or incomplete, when data are not well organized, and when an effective delivery system is not in place. A commonly overlooked feedback involves the data flow from GIS users back to the database. In a complete data cycle, the GIS user is a source for updates, changes to existing data and metadata, and derived datasets. In addition, end users can provide a wealth of feedback concerning specific data quality problems, as well as successful processes for using data, and thus provide essential input that can improve system design. Peer review of data is an important, often overlooked process that ensures reliable data. Highly reliable and well-documented spatial data are the lifeblood of an EGIS. Without a complete geospatial data cycle guiding the flow, data can quickly become unusable. Somers (2002) suggested that institutions should view spatial data as a "corporate asset" and protect it as such.

The integrity of the geospatial data cycle is the fundamental means to protect these institutional assets.

At the core of our enterprise design is the third key element, the spatial data warehouse, which ensures a complete geospatial data cycle and is tied to the needs of the institution. In these early stages in the evolution of EGIS, the focus tends to be data centric (von Meyer 1999). A spatial data warehouse provides fundamental framework for data management, the ability to stage, store, and deliver spatial data to a number of users in the institution. Smooth functioning of the data warehouse is enabled by a tracking system that follows both data and work flow. This tracking provides a conveniently available record of what processes have been completed and whether additional work is required. This method is analogous to "cradle to grave" tracking, but rather is "source to archive". In essence, the spatial data warehouse is a tangible implementation of a complete geospatial data cycle.

There are a number of strengths and advantages to our design and implementation for EGIS. Foremost among those advantages in design are: 1) the level of completeness in planning; 2) the achievement of a large degree of institutional, procedural, and technical interoperability with both logical and physical system components; and 3) the ability of the spatial data warehouse to stage, store, and share highly available and reliable data for all stakeholders while maintaining an unbroken geospatial data cycle that can support a broad spectrum of institutional GIS activities.

Comprehensive planning through resource evaluation allows the institution to take stock, recognize stakeholders, and determine how the existing diverse and distributed GIS resources can be better integrated, thereby increasing institutional interoperability and reducing redundancy to better serve the needs of the larger institution. The modular and interoperable design based on the five-step process also provides a better means, compared with existing GIS models (e.g., Coiner 1997 and Oppmann 1999), to integrate both logical and physical system components. Typically there has been less emphasis on the interoperability or technical design of system components and more emphasis on management of the system. Recognition of the many different interfaces and communication necessary between modular components is readily addressed in the framework of the five-step process, thereby producing a design with high interoperability. Procedural interoperability (Fletcher 1999) achieved through the use of policies, standards, and procedures provides a framework to integrate both the workflow and data flow within the enterprise.

In developing an EGIS it is advantageous to decrease the dependence on the corporate knowledge held by individuals, but rather to emphasize a functional system to track, sort, and stage data. Our logical system design allows quick location of original, staged, and processed data by any GIS analyst with appropriate permissions. Data that are well organized and documented tend to be more useful and widely used (Table 2: Utility of Spatial Data Warehouse). This usefulness is especially true in the regulatory environments where accurate records are required concerning the lineage of data, and where poorly organized and undocumented datasets can lead to costly litigation.

The process of measuring success of our EGIS transcends the completion of individual tasks, where emphasis is put on tangible elements and readily quantifiable criteria, such as system availability and usage (Table 2). However, the success of EGIS may be better measured by less tangible elements and harder to quantify criteria, such as elimination of data and resource redundancy, increased data sharing, increasing productivity, and increasing incorporation of GIS capabilities into the everyday work of the institution.

In developing the GISLab EGIS solution based on our five-step methodology, we met our strategic goals. We provided a highly reliable, available system that is used widely by our institution (Table 2). Our EGIS has heuristic value and general applicability, and, as such, it serves as a prototype for the form of an EGIS solution for the larger LANL complex. In a sense, GISLab blurred the boundary between a "project" or "departmental" GIS and an institution-wide EGIS. Although GISLab was established to serve two large projects (CGRP had a budget of $300 million, ER had a budget of $48 million), these projects spanned several divisions and many different institutions within the LANL complex. As such, the implementation of the GIS to serve these projects was by nature an enterprise solution.

At the close of the CGRP project in 2003, EGIS concepts are gradually being implemented at the level of the larger institution. GIS is being restructured at LANL. GIS professionals and their managers are grappling with EGIS issues: 1) better definitions of institutional GIS roles for site operations, project support, and basic R&D; 2) designation of institutional framework data; 3) development of an enterprise metadata clearinghouse; 4) establishment of a centralized spatial data warehouse; 5) establishment of off-site data backup; and 6) development of an institutional spatial information management plan. Hence, the process of implementing an EGIS at LANL has passed through a prototype phase and is now expanding to include the institution as a whole.

EGIS holds tremendous potential to aid in the decision-making process. A well-constructed enterprise provides highly available, reliable, and well-documented spatial data and analysis tools to decision makers -- the critical first layer in the foundation of a Spatial Decision Support System (SDSS). In addition, the integration of data within the EGIS, supplied by subject area experts, provides decision makers with a more complete picture of the spatial and temporal context for events during the decision making process; e.g., experimentally based information and data collected from the field can be integrated with remotely sensed data to evaluate likely scenarios. The predictive capability can be further enhanced when data are coupled with numerical models and simulations. Similar to a spatial data warehouse, a numerical model warehouse could store thousands of simulations from a variety of models, provide for query in the event that simulated model conditions approximate actual real-time conditions, and deliver results via the EGIS to decision makers.

Technology is the enabler for an EGIS and at times the driver, rapidly changing and shaping the products of GIS professionals. Technology is often seen as a "silver bullet" that solves management or technical challenges. However, it is limited and fallible, and requires institutions to fully ascertain the advantages of incorporating new technologies into their system design to maximize a positive impact. The evolution of EGIS is driven by technology and harnessed by R&D. An active R&D capability, side-by-side with routine operations, enables an institution to develop new technologies and improve processes, while meeting critical needs. While many of the components needed for EGIS are commercially available, the GIS software industry does not yet provide the "universal solution." Custom tools and novel design are required for most institutions. By responding to feedback from GIS stakeholders, along with technology transfer of broadly applicable custom tools, the commercial industry can provide increasingly complete enterprise solutions.

Stakeholder Consensus Tools: Stakeholders play vital roles in all aspects of EGIS design, development, and implementation. Stakeholder consensus-building tools and the enabling technologies of the intranet and Internet will be crucial in achieving the horizontal interoperability necessary to further build the bridges between EGIS at different scales, e.g., among local and regional institutions and stakeholders who wish to coordinate activities regarding large EGIS development. The Internet provides an integrating medium in which the collective concerns and opinions of large stakeholder groups can be synthesized and in which the resulting patterns may be understood. By making these patterns explicit and visible for the whole community, such that everyone has a better understanding of the global "lay of the land," better collective decisions can be made (Rasmussen et al. 2003).

The success of the consensus-building system developed to help guide the development of the CGRP GIS highlights the possibilities of this method. Given that the stakeholders in a planned EGIS may be geographically distributed, the Internet provides a means for involving them in the design process, not unlike public participation GIS efforts (Carver et al. 2001). A website provided basic information about the CGRP GIS effort and links to relevant portions of detailed databases and GIS maps. Through the use of online, open-response surveys, stakeholders identified the most important issues for CGRP-GIS, including how they were interrelated and how they related to the different stakeholders. The tool clarified potential conflicts among stakeholders by identifying who was in conflict with whom as well as why they were in conflict. Finally, it provided detailed documentation to support the final decisions on elements of the GIS design (Keating et al. 2002a). This collective intelligence method has been used successfully in the past to explore trends within a scientific community (Rasmussen et al. 2003), to aid strategic planning within a LANL science division, and to evaluate governmental efficiency.

Process Differentiation: Currently there is no process differentiation, or prioritization, of the numerous data and service requests sent through the intranet and Internet. Consequently, processes requiring real time access to the EGIS and those that do not are treated equally in terms of bandwidth and access priority. In the event of a disaster such as the Cerro Grande Fire, quality of service determinations would allow GIS applications to be treated differently allowing prioritization and reliability in the event of emergency (Mynard et al. 2003), e.g., emergency response workers would have the highest priority to access spatial data in the warehouse in the event of emergency. Differentiated service for the Internet is also needed to help keep pace with the evolving needs of multimedia GIS, real time video, and other real time traffic that require higher prioritization than simple data queries (Atiquzzaman 1999). The future EGIS will require new methods and technologies to overcome current service limitations.

. Net Developments: Enterprise-enabling software developments, such as Microsoft's .net Extensible Markup Language (XML) architecture promote good integration of information systems, devices, and client-to-client communication. XML-based application services allow universal information exchange across the intranet and Internet between otherwise disparate sources. Through XML application services, EGIS can expand beyond the server-client paradigm to a more interactive and inclusive one in which there are more client-to-client and service-to-service interactions than in current EGIS. Within an EGIS, XML-based application services may provide a better means for stakeholders to communicate data, and the processes or services performed on the data, to other stakeholders within and external to the EGIS. In many ways, the advent of the .net XML technology promotes the idea of "self describing" data as means to de-emphasize the dependence on a centralized EGIS to one that is more de-centralized, client-to-client data sharing.

If there are more client-to-client and less server-to-client interactions occurring within an institution what will the future EGIS look like? In many ways the .net XML technology stresses the importance of the key elements and flexibility needed in design. Stakeholder roles and the geospatial data cycle are perhaps more important in a de-centralized environment than a centralized one. It is more likely that technologies like .net XML will result in a hybrid system--neither fully centralized nor fully de-centralized.

While this paper has focused primarily on technical design, further work is needed to address management uncertainties, such as fiscal stability and viability of EGIS within an institution. One thing is certain: sound EGIS technical design is essential for responsible management of GIS data and applications.

6 Conclusion

Enterprise GIS is a phenomenon whose time has arrived. Now that key technological components are widely available – high-speed networks, fast computers, and sophisticated GIS analysis and visualization capabilities – we are ready for the next stage in the evolution of GIS, implementation of GIS at the enterprise level. As we have demonstrated, EGIS requires careful design, and must include careful consideration of stakeholder roles and needs, recognition of what is needed for a complete geospatial data cycle, and application of sound data warehouse design concepts. Our "five-step process" – design specification, resource evaluation, logical system design, physical system design, and implementation – leads to a sound technical and business design that serves institutional needs. The eventual result will be a new "collective geographic awareness", whereby both GIS specialists and non-specialists can access a wealth of map-based data and analysis capabilities, available twenty-four-hours a day, with minimal assistance, for the benefit of day-to-day operations, research, and responsible decision making.

7 Acknowledgments

This work was funded in part by the Cerro Grande Rehabilitation Project Office and the Earth and Environmental Sciences Division at Los Alamos National Laboratory. Thanks to Candace Batts, Justin Deming, Emily Hinz, Marcia Jones, Rick Kelley, Hari Khalsa, Steve Linger, Randy Mynard, Thomas Riggs, and Doug Walther for their valuable input.

8 References Cited

Atiquzzaman, M., 1999. Editorial special issue: enterprise networks. Computer Communications, (22), pp. 1281-1282.

Bernhardsen, T., 1999. Geographic Information Systems: An Introduction 2nd Edition. Wiley, New York, New York.

Bleakly, D., and Lee, R., 2001. Issues for consideration for long-term spatial data archiving. In Proceedings of 13th DOE Technology Information Exchange (TIE) Meeting, Albuquerque, NM November 13, 2001.

Bonham-Carter, G.F., 1994. Geographic information systems for geoscientists: Modeling with GIS. Pergamon/Elsevier Science, Kidlington, UK.

Brady, D., 2000. Designing GIS for high availability and high performance. In Proceedings of The Fourth International Conference / Exhibition on High Performance Computing in the Asia-Pacific Region, Beijing, China, May 14-17, 2000.

Brown, N., 1996. Industrial-Strength Management Strategies. IEEE Software, (13), pp. 94-103.

Calkins, H.W., and Tomlinson, R.F., 1977. Geographic Information Systems: Methods and Equipment for Land Use Planning. International Geographic Union Commission on Geographical Data Sensing and Processing, Resource and Land Investigations (RALI) Program, U.S. Geological Survey, Reston, Virginia.

Cardellini, V., Casalicchio, E., Colajanni, M., and Tucci, S., 2001. Mechanisms for quality of service in Web clusters. Computer Networks, (37), 761-771.

Caron, C. and Bedard, Y., 2002. Lessons learned from case studies on the implementation of geospatial information technologies. Journal of the Urban and Regional Information Systems Association, (14), pp. 17-36.

Carver, S., Evans, A., Kingston, R., and Turton, I., 2001. Public participation, GIS and cyberdemocracy: evaluating on-line spatial decision support systems. Environment and Planning B: Planning and Design, (28), pp. 907-921.

Choi, S.B., and Somani, A.K., 1996. Design and performance analysis of load-distributing fault-tolerant network. IEEE Transactions on Computers, (45), pp. 540-551.

Clinton, W.J., 1994. Coordinating geographic data acquisition and access: The national spatial data infrastructure, Executive Order 12906. Federal Register, (59), 17671-17674.

Coiner, Jerry C., 1997. Transferability of the Qatar Enterprise GIS Model: Experience in Vietnam and Jamaica. In Proceedings of GIS/GPS Conference '97, Doha, Qatar, March 2-4, 1997.

Dangermond, J., 2002, Personal Communication to P. Rich.

Dueker, K.J., and Butler, J.A., 2000. A geographic information system framework for transportation data sharing. Transportation Research Part C: Emerging Technologies, (8), pp. 13-36.

Federal Geographic Data Committee, 1998. Content standard for digital geospatial metadata. FGDC-STD-001-1998, http://www.fgdc.gov/standards/standards.html.

Fletcher, D.R., 1999. The Interoperable Enterprise. In Enterprise GIS. URISA, Park Ridge, Illinois, pp. 7-16.

Gan, X., Schroeder, T., Goddard, S., and Ramamurthy, B., 2000. LSMAC vs. LSNAT: Scalable cluster-based Web servers. Cluster Computing, (3), pp. 175-185.

Goodchild, M.F., 1995. Sharing Imperfect Data. In Sharing Geographic Information, edited by H.J. Onsrud and G. Rushton. Center for Urban Policy Research, Rutgers, New Brunswick, New Jersey, pp. 413-425.

Kirkpatrick, S.A., Wofford, J.C., and Baum, J.R., 2002. Measuring motive imagery contained in the vision statement. The Leadership Quarterly, (13), 139-150.

Keating, G., Rasmussen, S., and Raven, M., 2002a. Consensus building tools for GIS design. Los Alamos National Laboratory Report LA-13894-MS, Los Alamos, NM.

Keating, G., Rich, P.M., Witkowski, M.S., Batts, C.M., Deming, J.H., Jones, M.A., Linger, S.P., Mynard, C.R., Riggs, T.L., and Walther, D., 2002b. Challenges for EGIS in post-wildfire hazard mitigation and emergency management. Los Alamos National Laboratory Report LA-13930, Los Alamos, NM.

Keating, G., Rich, P.M., and Witkowski, M.S., 2003. Challenges for Enterprise GIS. URISA Journal, in press.

Lembo, A.J. Jr., 1999. Implementing Enterprise-Wide GIS for Government: An Example from Nassau County, New York. In Enterprise GIS. URISA, Park Ridge, Illinois, pp. 62-79.

Linger, S.P., Batts, C.M., Rich, P.M., Riggs, T.L., Walther, D.E., and Witkowski, M.S., 2002a. GISLab Enterprise GIS Tools. Los Alamos National Laboratory Report LA-UR-02-4008, Los Alamos, NM.

Linger, S.P., Walther, D., and Jones, M.A., 2002b. Rapid Mapping Application: An ArcMap tool using VBA, templates, and XML. In Proceedings of 22nd Annual ESRI International User Conference, San Diego, CA, July 8-12, 2002.

Lloyd, W.A., 2000. Beyond Enterprise GIS. In Proceedings of Twentieth Annual ESRI International User Conference, San Diego, CA, June 26-30, 2000.

Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W., 2001. Geographic information systems and science. Wiley, New York, New York.

Mynard, C.R., Keating, G.N., Rich, P.M., and Bleakly, D.R., 2003. GIS emergency support for the May 2000 Cerro Grande wildfire. Los Alamos National Laboratory Report LA-UR-02-6746, Los Alamos, NM.

Office of Management and Budget, 2002. Coordination of geographic information and related spatial data activities. Circular Number A-16, Revised 19 August 2002, http://www.whitehouse.gov/omb/circulars/a016/a016_rev.html#2b2.

Oppmann, R.S., 1999. An Organizational Model for Successful Enterprise GIS. In Enterprise GIS. URISA, Park Ridge, Illinois, pp. 44-52.

Peng, Z.-R., Groff, J.N., and Dueker, K.J., 1998. An Enterprise GIS database design for agency-wide transit applications. Journal of the Urban and Regional Information Systems Association, (10), 46-55.

Pinto, J.K., and Onsrud, H.J., 1995. Sharing geographic information across organizational boundaries: a research framework. In Sharing Geographic Information, edited by H.J. Onsrud and G. Rushton. Center for Urban Policy Research, Rutgers, New Brunswick, New Jersey, pp. 44-64.

Rasmussen, S., Raven, M., Keating, G.N., and Bedau, M., 2003. Collective intelligence of the Artificial Life community: Successes, failures, and the future. Artificial Life Journal, in press.

Rich, S., Das, A., and Kroot, C., 2001. Spatial data management in an Enterprise GIS. In Proceedings of Twenty-first Annual ESRI International User Conference, San Diego, CA, July 9-13, 2001.

Rush, S.F., 2001. Establishing a framework to improve geospatial data management at Hanford. In Proceedings of 13th DOE Technical Information Exchange (TIE) Meeting, Albuquerque, NM, November 13, 2001.

Somers, R., 2002. Managing your move to Enterprise GIS. Geospatial Solutions, (12), 22-26.

von Meyer, N.R., 1999. Operational Data and the Data Warehouse. In Enterprise GIS. URISA, Park Ridge, Illinois, pp. 53-61.

Witkowski, M.S., Rich, P.M., Keating, G.N., Linger, S.P., and Riggs, T.L., 2002. Spatial Data Warehouse Design for Enterprise GIS. In Proceedings of GIScience 2002, Second International Conference on Geographic Information Science, Boulder, Colorado, September 25-28, 2002.

Xiong, D., and Lin, H., 2000. Spatial Data Handling for ITS: Perspective, Issues and Approaches. GeoInformatica, (4), pp. 215-230.