Geospatial Data Repository

by Marilyn O. Ruiz

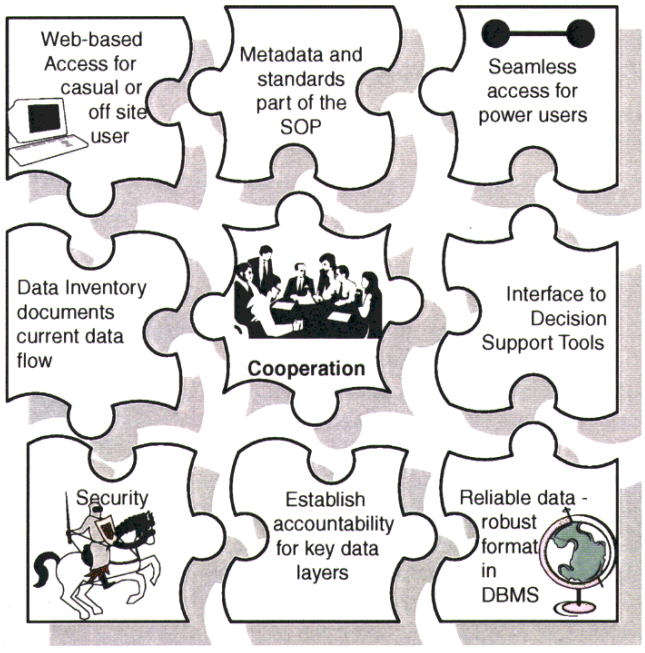
ERDC/CERL TN-01-1 February 2001

Sharing Data Across the Organization and Beyond

Problem

Historically, the organizational framework for geospatial data has been site- and software-specific. The complexity, size, and specialized use of geospatial data have been real impediments to integration into the mainstream of data storage and data delivery technologies. This is changing. Advances in database management systems (DBMS), object-oriented programming, and data models, and improved data delivery via the Internet and Intranet open up significant new avenues for improved management and delivery of GIS data.

Certain challenges remain. To take full advantage of the new options, business processes must be re-evaluated and often revised. Technology decisions require considerable understanding of a number of rapidly changing software products. Research at the U.S. Army Engineer Research and Development Center (ERDC) has focused on the development of strategies and solutions for the architecture and implementation of data repositories for land management, taking into account recently developed technologies for this purpose.



In the puzzle of sharing geospatial data, cooperation and communication are key to success.

Key Issues

Duplication of data themes in various versions across offices and computers.

Uncertainty about accountability for the content of key data themes.

Turn over of GIS staff with a subsequent loss of institutional knowledge.

Data requests taking time away from other duties.

Security concerns over unintentional release or misuse of sensitive data.

Approach

A five-part approach was developed to create a repository that addressed the key issues identified.

Identify Repository Stakeholders. The stakeholders include those who can influence the repository in some way, can benefit from it, have invested resources in it, have an interest in the outcome, or have other programs that may depend on the effectiveness of the repository.

Data Inventory. The data inventory is used to collect information in a systematic way across the servers and the workstations of the primary users. A data survey form assists in the collection of information, which is then transferred to a database file. Information collected includes data content, update frequency, spatial extent, data source, and quality or security concerns.

Data Storage, Search, Upload, and Download. Data storage, search, upload, and download are the primary operations required for the repository. Technology and managment options need to be evaluated and then integrated to develop an operational system. A report on the review of technology and development of a prototype is found in a forthcoming ERDC technical report.

Web-based Portal. A web-based tool was developed that facilitates search, download, and upload functions by setting up a standard manner in which to handle metadata, by indexing the data in the repository, and by providing a framework for data administration.

Security Assessment. A professionally executed security assessment provides guidance for identification of and remedies for potential security threats. These would include misuse of the data in the repository through both negligence or malicious intent.

Geospatial Data Repository

Fort Hood, Texas

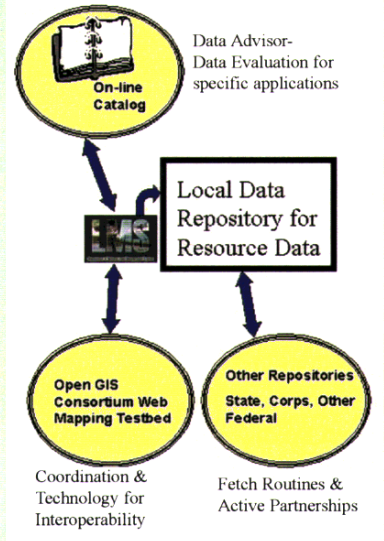
Fort Hood, Texas, was the site for development of a prototype system that supplies a more centralized GIS solution. Fort Hood is a Land Managment System (LMS) field demonstration site, and the repository was designed to be compatible with the modeling application goals of LMS as well as with those of the Fort Hood stakeholders. At Fort Hood, GIS technology has been used extensively for military land management. Though mature in many ways, there remained some issues related to geospatial data managment that hindered efficiency. The Fort Hood project was called the Data Enterprise Repository (DER). Participants in the DER included:

The Environmental Division's Natural Resource Branch (NRB) and Cultural Resource Team.

The Integrated Training Area Managment office.

The Nature Conservancy of Texas, which has an office on-site through contract with the NRB.

The most important source of data for the repository was the data used by Fort Hood personnel. The most up-to-date and well-conceived data about a place is usually produced by people at that place. The challenge is to create a management and technology system that makes that data available on demand.



Characteristics

Tools for data discovery, access, and archiving

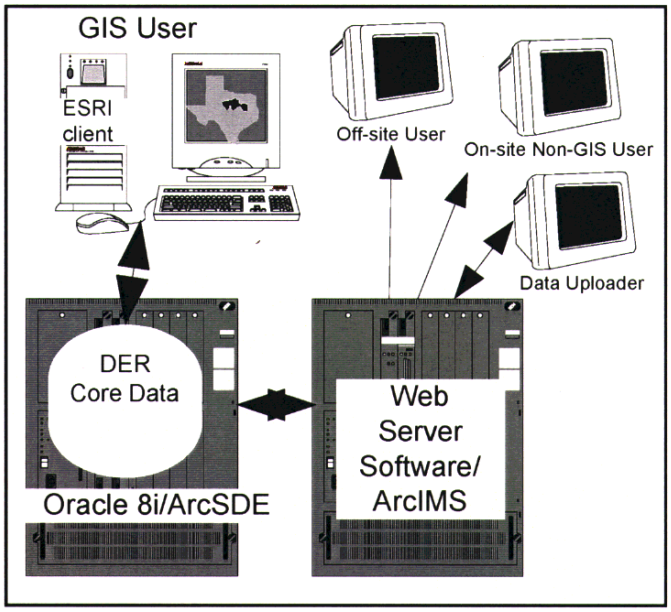
Seamless, web-based system built on a foundation of Oracle 8i coupled with ArcSDE and ArcIMS

Standards for data formats, data models, and metadata

Decision Support at the local level involves local data. Much of the data needed by installation personnel may be in a centralized repository. Other data comes from other agencies and repositories. A combination of fetch routines along with interagency cooperation and partnerships helps to ensure that the data used is current and reduces data redundancy. The LMS Model Catalog provides metadata about models.

Solution

The geospatial data repository design aimed to address three main areas of concern dealing with corporate knowledge, site-specific data management, and a coherent data delivery environment. The technology behind the solution is based on Oracle and the ESRI product suite with a focus on Arc8, Arc Spatial Database Engine (ArcSDE), and Arc Internet Map Server (ArcIMS).



Simplified Diagram of the Hardware and Software for the Prototype Enterprise System.

Design Considerations

Corporate Knowledge. It was found that the installation staff that uses data daily is generally well-informed about the location and content of geospatial data at the installation. Some of that knowledge could be shared beyond the core group or specific individuals with a more standardized method of documentation. The current method of data sharing requires tapping into the knowledge of key individuals. For those people, the need to respond to requests for data can be an impediment to fulfilling other critical functions. The method does not function when personnel are absent and it breaks down with turnover in staff.

Site-Specific Data Management Issues. While many data management issues are common to many places, the solutions need to be site specific. Lack of accountability for the content of data layers is a common issue, for example. The manner in which individuals are held accountable will be most successful when there is agreement about the responsibility and a method is devised that allows for fulfillment of that responsibility. This is true of many mangament issues.

Geospatial Data Repository

Coherent Data Delivery. The multiple complexities of data volume, GIS data formats, multiple shared servers, and physical distance among data users have been impediments to seamless sharing of geospatial information and related databases. All shared data needs to hava a coherent "front end" so users see a common, well-documented database. A more automated, coherent method of data access alleviates the pressure on individuals by standardizing corporate knowledge and provides a more efficient and effective way to share data.

The Role of Metadata and Standards

Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata must be provided for all geospatial data. This is necessary for re-use and to allow an evaluation of quality of data. See http://www.fgdc.gov/metadata/metadata.html

Dublin Core Initiative metdata provides a small set of widely used core information to identify key attributes of each data layer and is especially helpful for data search. See http://purl.oclc.org/dc/

Spatial Data Standards (SDS) for Facilities, Infrastructure, and Environment was developed by the Department of Defense CADD/GIS Technology Center. The SDS provides a comprehensive framework for the design of a geospatial database. In the SDS all data fall under one of the Standard's 26 entity sets, such as flora, geology, or transportation. Each entity set is further subcategorized into classes, which provides for the additional clarification of data content. All of the data in the repository are referenced by their SDS entity set and class. See http://tsc.wes.army.mil/products/TSSDS-SFMS/tssds/html/

Technology Issues and Developments

The enterprise system needs to provide the main functions of data storage, discovery, upload, and download from several perspectives. In particular, the data need to be available with:

an application programming interface for linking to custom applications

seamless access through a GIS

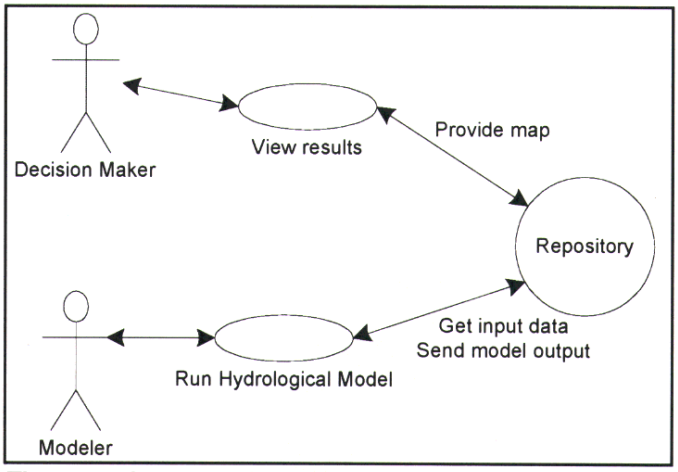
web-based access without the need for a GIS.

Access for Applications. For applications, a geospatial database serves several functions. It provides (1) spatial location and attribute information required by the application as input, (2) direct access to map visualization options to better understand application results, and (3) a database storage option for data the application outputs. Examples of applications include environmental models such those predicting soil erosion or wildlife populations as well as management tools that help to automate the daily work related to land management.

ArcSDE provides the primary outlet from the prototype system to applications. This is a benefit to application developers, because it provides "one stop shopping for data." ArcSDE is also needed to access the geospatial information when that data is stored using the ESRI geodatabase approach. The geodatabase is an object-oriented data model that allows the user to add "intelligence" or "behavior" to the data.

For example, it is common to store the boundary of an installation as one theme and the training areas of an installation as a separate theme. In a geodatabase, these two themes could be linked, so that a change in the boundary file would be reflected directly in the training area map. In addition, allowable names for training areas could be entered as a "domain," helping with quality control.

Since the attribute and spatial data are stored in Oracle tables, Oracle database tools may also be used to access the attribute information. The spatial data is accessible through Oracle if it is stored in the Oracle Spatial format. ArcSDE reads and creates Oracle Spatial, but geodatabase characteristics are not available with this method.

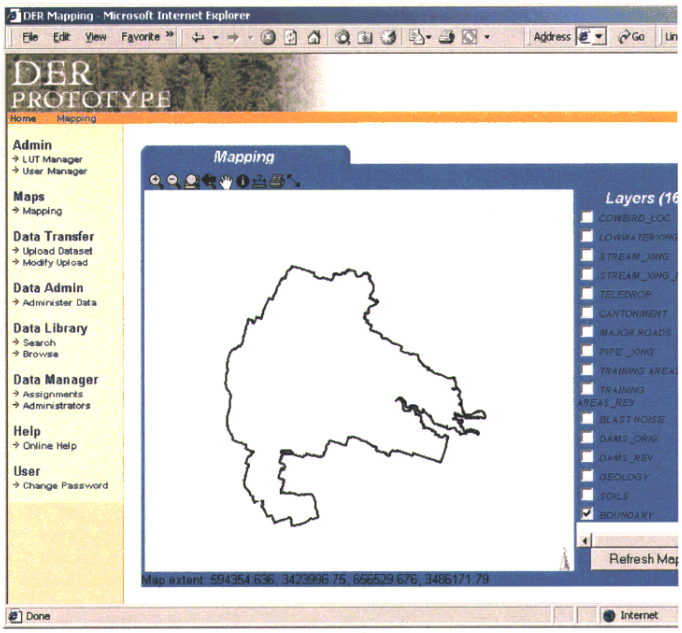


The repository provides input data, links to visualization tools, and a storage option for data output from applications.

Geospatial Data Repository

Access Through GIS or CADD. For core GIS users, the ESRI Arc8 software tools link directly to the repository. A CADD option in ArcSDE also allows CADD access. If the Oracle Spatial option is used for spatial data, other GIS or CADD programs can link to the data as well. Once the link is made, the data in the DER are simply part of the data available in the GIS or CADD environment. The data may be altered only by designated users.

Web-Based Access. For non-GIS users, for users off site, and for users who need to contribute data to the repository, a web-based application provides a full set of options. Users can search for data using the SDS fields, or the Dublin Core Metadata. Once selected, maps of the data may be displayed or printed. A data upload component facilitates the contribution of material to the repository along with appropriate metadata. Following data upload, a work flow manager helps the data manager to index the data in a standard way. This ensures that all data has a level of quality control, that metadata is consistent, and the searchable fields are robust. The data download option provides several common geospatial data formats, which are prepared, on request, directly from the core data in the repository.



Direct display and printing of maps are some options available through the web-based repository application.

Benefits

An enterprise geospatial data repository improves the efficiency and effectiveness of geospatial data management at the local level. Specific benefits include:

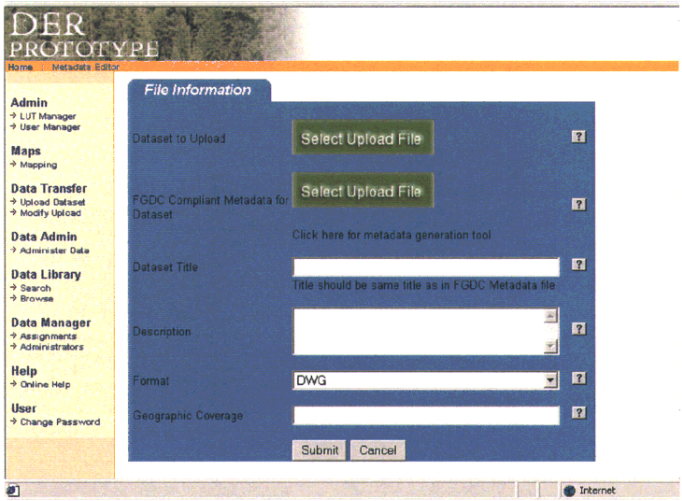
In cases where GIS activities are dispersed across an organization, the repository serves as a common focal point for data storage and access.

Compliance with the requirement to supply metadata is made part of the business process. This reduces the potential for liability related to use of incorrect or uncertain data.

GIS analysts can spend more time providing decision support assistance and less time on meeting requests for data.

With a common, agreed upon framework for data storage, upload, and download, less staff time is spent on designing solutions *ad hoc*.

Connections to applications is more straightforward in a system based on a DBMS than in a system based on data read from files directly.



Data upload starts with submission of data and metadata.

For More Information

For information about the Fort Hood Prototype Data Enterprise Repository or on-going research on geospatial data management at military installations, contact:

Marilyn O. Ruiz, PhD, Principal Investigator
U.S. Army Engineer Research and Development Center
Phone: (217) 352-6511 x7368
Fax: (217) 373-7251
E-mail: Marilyn.Ruiz@erdc.usace.army.mil

Find Geospatial Data Repository Information on DENIX under the LMS program at:

http://www.denix.osd.mil/denix/Public/Library/LMS/lms.html