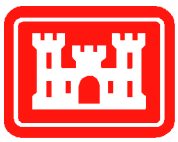
ERDC TR-00-9



US Army Corps of Engineers®

U.S. Army Corps of Engineers Geospatial Data and Systems Management

Nancy H. Greeley, Kelly M. Dilks, Chad M. Adams, and Gregg C. Hoge

November 2000

Engineer Research and Development Center

Approved for public release; distribution is unlimited.

Abstract: To help the U.S. Army Corps of Engineers manage its substantial investment in geospatial data and systems (GD&S), a review of the current state of GD&S in the Districts and Divisions was needed. The authors surveyed employees involved in GD&S work in multiple functional areas at multiple sites. The survey contained questions on GD&S administration and personnel issues including data management, GD&S maintenance, data libraries, data sharing, and Information Management involvement with GD&S. Other topics areas covered are GD&S workflow and technical issues, including hardware, software, metadata creation and publishing, data storage and distribution, networks, and security. The results of the survey show great differences in the approaches and challenges within the various Districts and Divisions. Many are finding solutions that are shared in this report. The report is written as a manual for managers. Each existing problem is described followed by possible solutions. Seven conclusions and a summary "Manager's Checklist" of GD&S issues are included.

How to get copies of ERDC technical publications:

Department of Defense personnel and contractors may order reports through the Defense Technical Information Center:

DTIC-BR SUITE 0944
8725 JOHN J KINGMAN RD
FT BELVOIR VA 22060-6218
Telephone (800) 225-3842
E-mail help@dtic.mil
msorders@dtic.mil
WWW http://www.dtic.mil/

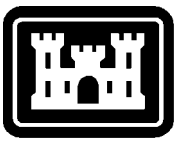
All others may order reports through the National Technical Information Service:

NTIS
5285 PORT ROYAL RD
SPRINGFIELD VA 22161
Telephone (703) 487-4650
(703) 487-4639 (TDD for the hearing-impaired)
E-mail orders@ntis.fedworld.gov
WWW http://www.ntis.gov/index.html

For information on all aspects of the Engineer Research and Development Center, visit our World Wide Web site:

http://www.erdc.usace.army.mil

Technical Report
ERDC TR-00-9



US Army Corps
of Engineers®
Engineer Research and
Development Center

U.S. Army Corps of Engineers Geospatial Data and Systems Management

Nancy H. Greeley, Kelly M. Dilks, Chad M. Adams, and Gregg C. Hoge November 2000

Prepared for
OFFICE OF THE CHIEF OF ENGINEERS

Approved for public release; distribution is unlimited.

PREFACE

This report was prepared by Nancy H. Greeley, Physical Scientist, U.S. Army Corps of Engineers, Engineer Research and Development Center (ERDC), Information Technology Laboratory (ITL), Hanover, NH; Kelly Dilks, Geographer, ERDC, Construction Engineering Research Laboratory (CERL), Champaign, IL; Chad M. Adams, Senior Database Programmer, Systems Engineering and Security, Inc., Hanover, NH; and Gregg C. Hoge, Chief, ERDC Information Technology Coordination Center, ITL, Hanover, NH.

Kelly Dilks was the lead investigator and was responsible for initiating the program and presenting the progress to the Field Review Group. The team relied heavily on Ms. Dilks's vast knowledge of Districts and Divisions. The CERL and ITL team worked in partnership to devise the data collection strategy, collect and analyze the data, and produce the report.

The authors thank all of the District representatives who were interviewed. They thank Nancy Blyler, Headquarters, Corps of Engineers, Civil Works, who encouraged and guided them throughout the entire process and reviewed the report. They thank Rory Sutton, Corps of Engineers, South Atlantic Division, Jacksonville, FL, who reviewed the report and wrote Appendix E. Finally, they appreciate the work of Dr. Chris Rewerts, CERL, and Tari Weicherding, Department of Geography, University of Illinois, who wrote Appendices D and F, respectively.

This report was prepared as the final product of the FY 98 U.S. Army Corps of Engineers Directorate of Civil Works Program, Work Unit Number 33172, *Geospatial Data Management Infrastructure: Guidance to Managing USACE Data.* It will be available via the USACE GIS web site (http://gis.usace.army.mil).

The contents of the report are not to be used for advertising or promotional purposes. Citation of brand names does not constitute an official endorsement or approval of the use of such commercial products.

CONTENTS

Preface ii
Introduction 1
Purpose 1
Method 2
Results 2
USACE geospatial data and systems management: Administration and personnel issues 2
Geospatial data and systems maintenance 4
Geospatial data standards 5
Geospatial data library 7
Data sharing 7
Optimizing geospatial data work flow and document flow 8
USACE geospatial data and systems management: Technical aspects 10
Data storage media 12
Network architecture 13
System security 15
Conclusions 15
Literature cited 15
Web resources 16
Appendix A: USACE geospatial data and systems management: Manager's checklist 17
Appendix B: Sample hardware configurations and procedural recommendations 23
Appendix C: Geospatial data management survey 25
Appendix D: Spatial data engine prototype for the Mississippi Valley Division 31
Appendix E: Jacksonville District: data archiving, data management, and continuity of operations in the geospatial data arena 35
Appendix F: Related work: the Alexandria digital library project and on-line access to information via geographical references 39
Appendix G: USACE sample network architecture and sample organization network architecture 43
Abstract 45

U.S. Army Corps of Engineers Geospatial Data and Systems Management

NANCY H. GREELEY, KELLY M. DILKS, CHAD M. ADAMS, AND GREGG C. HOGE

INTRODUCTION

The US Army Corps of Engineers (USACE) is in the business of geospatial data creation, yet we do not have a corporate plan for consistent implementation and use of Geospatial Data and Systems (GD&S) throughout the organization. Headquarters and each District, Division, laboratory, center, and field operating office has different functions, funding sources, personnel, and technology environments. The responsibility for GD&S at each District and Division is left to local managers. The result is that the GD&S environments at some District and Divisions are productive and forward-looking but at others are just passable and reactive.

An example of differences within USACE's GD&S community can be seen in how they have dealt with the task of geospatial metadata creation. Since 1994, USACE employees have been required to create geospatial metadata corresponding to their geospatial data files (Executive Order 12909, USACE ER 1110-1-8156, and USACE EM 1110-1-2909). Some Districts have embraced the directive and routinely create metadata for all new data files. However, many USACE employees don't know what geospatial metadata is, how to create it, or what to do with it when it is created. Many managers don't know how to fund creation of the required metadata.

Managing geospatial data and geospatial data systems requires a high degree of technological sophistication. Yet because USACE commands do not have a centralized GD&S office, there is no one proponent for GD&S at any given site. There is no one agent assigned to manage the databases, or the required software or hardware, to report up the chain of command on local successes, or to champion funding needs. At some commands the local Information Management (IM) specialists work closely with those using geographic information systems (GIS) or computer-aided drafting and design (CADD) systems, and at others, IM personnel are not involved at all.

Individual managers and employees with the willingness, ingenuity, and foresight to see the potential value of GD&S and with the resources and abilities to implement it have often been successful. However, there are many hurdles to successful use and management of GD&S including decreasing funds, increasing technological requirements for hardware and software, exponentially increasing amounts of geospatial data, and increasing expectations of direct-funding customers and the public.

USACE must be competitive with other federal agencies in the geospatial data arena or we will lose business to those who maintain their expertise. We must serve our customers and the public the way they now expect to be served in this arena. To do all this, USACE managers need the resources to empower their employees to develop, maintain, and use the highly sophisticated GD&S technology. Managers of the separate stovepipes within each command that can offer something to and need something from GD&S need to work together, looking at ways to optimize the GD&S within their organization. Leaders of the managers must be sure they are doing so.

PURPOSE

The purpose of this report is to

- Identify strengths and weaknesses in the way that USACE manages its GD&S.
- Investigate alternatives and recommend changes to correct weak management strategies.
- Offer descriptions of stronger management strategies in USACE experience so that managers with a GD&S aspect in their programs can determine if these strategies could work in their GD&S groups.

We hope that this report will be used as a resource by USACE leadership, managers, and staff involved in GD&S, enabling them to take a new look at their GD&S capacities and options. Ideally, it will be the springboard for greater communication within the USACE GD&S community.

METHOD

A four-pronged approach was used to conduct this research:

- Visits to and surveys of USACE Districts and Divisions
- Funding and monitoring of GD&S management pilot tests in several locations
- Review of various GD&S implementation plans and reports
- Investigation of GD&S management practices in other government agencies and private industry.

RESULTS

What follows is a discussion of GD&S management in USACE, some problems and successes in the GD&S arena that were identified during this study, and recommended solutions.

USACE geospatial data and systems management: Administration and personnel issues

Geospatial data management

During implementation of geographic information systems (GIS), data management was often ignored. This was in part because the amount of data available for use with GIS was relatively small in size and often fit on one UNIX host computer. Technicians did not find issue with the data management aspects of GIS implementation. Today, the amount of data available is counted in gigabytes instead of kilobytes and is stored in a myriad of locations. While conducting this project, researchers asked USACE personnel what they felt were the biggest data management issues from their perspective. The following describes the issues raised from the ensuing discussion.

Problem: Data discovery

Description

The most prevalent problem related to geospatial data management is the difficulty in finding the data necessary for a project. Many Districts have one or two knowledgeable users who are aware of the majority of the District's holdings. These individuals are always in demand and spend much of their time helping others, falling behind on their own regular duties. Because data discovery is not identified as "real work," it is seen as robbing time from more traditional tasks. If the knowledgeable user is on travel duty or leave, backup data discovery mechanisms often do not exist, causing unnecessary delays in the project.

Solutions

To enhance data discovery, we recommend the approach some Districts are taking in testing out various commercially available Web mapping software applications. Automation of data discovery will initially require more time and expenditure than maintaining the status quo, but the investment will bring a good return in saved time and resources in the future.

Software being tested includes ESRI's Arc Explorer and Internet Map Server, Bentley's Geomap, and Intergraph's Geomedia products. These products allow users to browse and access available data graphically. Systems are optimized when users do not maintain separate copies of the same data and when data are stored in a virtually centralized location. (These issues are addressed in later sections of this report.) This type of intranet Web viewing of data assumes that the local area network (LAN) has sufficient capabilities to handle the size of these data files. Some GIS data files are several gigabytes in size, and these software products may identify deficiencies in the LAN.

One District is using the technology described above in conjunction with hiring a geospatial data manager. The geospatial data manager can act as a conduit of information related to the acquisition of data, the use of data, the metadata related to the data, and as a digital data librarian to ensure prudent measures are taken to lengthen the life cycle of the data. Even after the decision was made to fund this position on a part-time basis in a District, finding a person with the required knowledge and abilities, and a manager willing to give up some portion of this employee's time, was difficult.

Problem: Time and money

Description

Time and money are constant constraints in the USACE environment. GIS requires a commitment from management of both time and money. GIS implementation is not a one-time procurement with start-up costs, but a life-cycle system. People often plan for hardware and software maintenance but ignore the data maintenance requirements.

Solutions

One District was particularly proactive as to their GIS goals. They drafted a GIS implementation plan and had all managers sign off on it and commit to providing a percentage of funds for its use in the future. The amount of funds required was determined by the project acreage, so the project manager could estimate

the required resources. This was successful because the lines of communication were open and management supported integrating new technologies into the District's workflow.

Problem: Understanding the power of GIS

Description

A common issue in the USACE related to GIS is the lack of understanding about what GIS can do for the organization. Many USACE managers see it as another duty instead of as a tool for decision making. This point of view inhibits use of GIS in support of the USACE mission. GIS enables workers to make sound and supportable decisions for their projects. The time involved in the decision process, and thus the cost involved, is not increased, but usually decreased by the use of technology and GIS tools.

Solutions

For Districts just starting down the GIS path, it is suggested that interested individuals get copies of other Districts' GD&S or GIS implementation plans. Ideal champions of the effort are the Executive Office, for GIS use, and the IM community, for the management of the data. The role of GIS in support of the civil, military, and environmental missions of the USACE is expanding. GIS is taught to cadets at West Point, which shows the Army's commitment to using appropriate technology in support of the mission. Most state, local, and federal agencies are using or plan to use GIS, and they expect the USACE to use it in projects in which they partner with USACE.

Problem: Communication and coordination

Description

We found that there is no clear definition of which groups within each District or Division are responsible for the creation and maintenance of geospatial data or for the maintenance of GD&S hardware and software. In each District and Division, the first group in the organization that needed to use GD&S or that realized that GD&S would be a valuable tool for the organization was the one to start developing GD&S capabilities. In some cases, after this initial buy-in by one group, other groups in the organization refused to take any responsiblity for GD&S even though their cooperation would have made the GD&S more efficient.

Solutions

GIS is a technology that would benefit from a breakdown of the stovepipe nature of USACE. Data, applications, and analysis should be shared at the District level between sections, branches, and teams. This mentality must be encouraged by all managers involved. In groups with highly successful GD&S, communication between the groups often begins at the District GD&S technical committee meetings, as described in EM 1110-1-2909. These meetings are fundamental to the success of GIS implementation. Inclusion in the technical committee meetings of all divisions, branches, or offices that have some part in geospatial data production, storage, or use is ideal.

Problem: Responsibility for GD&S

Description

The most common home of GIS at USACE Districts is in the Planning Division. The historic home of CADD is Engineering. In some Districts, organizational elements are engaged in battles to control GIS. This infighting is counterproductive to the USACE mission and leads to misconceptions about GIS, which is a tool to be used for intelligent problem-solving. In smaller Districts, GIS funding is so sporadic that the responsibility for incorporating any GIS into the normal project process randomly falls onto whichever individuals are willing to work at it.

Solutions

In Districts with evolved GIS programs, there is usually more than one person responsible for GIS. These Districts pointed out that the process of evolution may have been further assisted by having one individual as the champion or lead for implementation. This GIS leader's responsibilities could include ensuring proper use of the system, providing application assistance to new users, engaging management in realizing the power of GIS, and marketing the District's GIS capabilities to potential new customers. However, accomplishing these tasks would require funding, and the costs may be more than some Districts are willing to pay.

When sections that historically have not used GIS want to become part of the program, they must be responsible for budgeting for the hardware, software, and assistance necessary to make it successful. In some successful cases, personnel have been detailed to another section of the District, either to learn GIS or to teach it to new users. Development of GIS capabilities is the responsibility of all managers involved, and the partnership must have commitment, both in terms of money and time for the employees, for it to succeed.

Problem: Contracting geospatial data work

Description

The contracting officer sees USACE language related to data and standards, but does not know how to interpret it.

Solutions

Communication is key when contracting out GIS services. There needs to be a closer relationship between the contracting office (CO) and the contracting officer's technical representative (COTR). The COTR's responsibility is to monitor procurements and deliverables at the technical level to ensure that the USACE is getting what it needs and expects from the contractor. For every geospatial dataset created, the deliverable must include a corresponding geospatial metadata file that conforms to standards set by the Federal Geographic Data Committee.

Geospatial data and systems maintenance

Data maintenance refers to the updating of a data collection within a time frame that is meaningful for the data type or for the project purpose. For example, data maintenance on river bottom cross-sections taken for navigation purposes requires much more frequent updates than a dataset of county-wide terrain elevation for a mapping project. Data maintenance also requires backup of datasets and maintenance of the hardware and software necessary to create and read data files.

USACE generally does not provide uniform guidance related to maintenance of geospatial data. This flexibility in requirements for updating is good considering the differences in the projects in each District or Division. A greater awareness of systems maintenance is sorely needed, however. This is discussed in the problem section below.

Data layers that are actively maintained vary by District. In some Districts, landcover may be updated every decade. In Districts with strong environmental restoration responsibilities, landcover is updated on an annual basis. Data collection update cycles can range from 15-minute cycles for automated systems to 10-year cycles for less important or less dynamic data layers.

USACE engineer manuals sometimes recommend that mission-specific data be updated at regular intervals. Based on these recommendations, some mission-specific data is actively maintained by Divisions, including

- Channel improvement master plan data
- Monitoring data such as groundwater, surface water, rainfall, and salinity
- Geodetic control data
- Navigation-related data.

Because maintenance schedules for geospatial data are so diverse, it would not be useful to create broad-based guidance on how often to update various themes of data. It would be useful to offer guidance related to the dos and don'ts of data maintenance scheduling. This would allow for the flexibility necessary to support all USACE projects and mission areas.

The survey found the following problems in the area of geospatial data maintenance.

Problem: Budgeting for data maintenance

Description

Data maintenance is often ignored in project budgets. USACE-owned data is usually kept for historical purposes, but data owned by other state or federal agencies may be thrown away if disk space becomes a constraint. When an outside source makes a data request from a District or Division, funds are not available to make updates, so the data is sent out 'as is.'

Solutions

Data maintenance must be put into project budgets for mission-required data. Overall, most respondents said that data maintenance is needed, but funding is not available so maintenance is done on an ad hoc basis. This is based on a corporate culture that is oriented to the short-term and doesn't look at data as a long-term investment. Labor for updates must be charged to projects and must be included in the project budget.

Problem: Data backups

Description

A wide variety of routines exist for geospatial data backups. Some locations do daily backups for all changes that occur and full backups once a week. Other locations do backups when they have a chance. The situation often varied within a District between CADD and GIS data backups and who was responsible for the data. In some instances, the backup schedule is increased when weather conditions indicate the potential for a natural disaster.

Solutions

The best backup and storage schema are at locations where the relationship between the geospatial data users and information management people is strong. Backup and storage recommendations can be found in the technical section of this report. There is absolutely no valid reason or excuse for not having a consistent, logical backup and storage plan. Each location paid for the original collection of data, and this investment must be protected.

Problem: Software and hardware updates

Description

As was seen previously in the data management

overview section, money is a driving factor in the timeliness of software and hardware updates.

Solutions

In the best situations, hardware, software, and data updates are incorporated into the various projects' budgets. This allows for appropriate technology to be used for the project. This provides a larger potential return on investment and potential bulk purchase discounts for maintenance and other updates.

There are some aspects of this dilemma for which no one seems to have found a satisfactory answer. Some USACE locations order enough software updates for every license but wait for the individuals to request that the software be put on their machines. This takes into account individual project milestones and does not take computers out of commission at a critical time in a project. Scheduling around projects may work well within the project timelines, but it facilitates multiple versions of the same software being utilized for the same project. At times, different versions of the same software inhibit sharing of data because of format changes or the addition of new data features.

Several Districts update their software all at once to avoid trouble with various versions. This maintains compatibility between the data and the system, but it may be scheduled at times that interfere with project deadlines.

In a few instances, software updates were only made when expected by a sponsor. Hardware was updated only when the software could not be run on the machine because of file size, memory requirements, or other technical requirements. This crisis management mentality does not allow GIS to be utilized to its fullest extent and may hinder potential new projects because USACE will not be able to hit the ground running on a project for timely completion.

Geospatial data standards

Coordinated data and metadata standards are necessary so data creators and data users from different organizations, fields of study, locations, and periods of time can communicate with one another. If a data creator uses definitions and techniques of measuring data that are not documented and have no relevance to the definitions or techniques being used by others, then his product will have no meaning for anyone other than himself and his own short-term project. The data could not be compared with the results of others, and there could be no long-term value or future use of the data.

We in USACE, with our long historical contribution to the nation and to science, must realize the long-term value of our work for the future and for use by other projects besides our own. We must make that possible by cooperatively using and contributing to the development of data and metadata standards.

Problem: Data dictionaries

Description

The CADD/GIS Technology Center was formed to support tri-service entities related to CADD and GIS standards. One product from the Center is the Special Data Standards for Facilities, Infrastructure and the Environment (SDSFIE). Since the SDSFIE has excellent potential as a standardized data dictionary for USACE GIS users, survey respondents were asked about their experience using this product.

One of the major problems, as seen by the field, is the SDSFIE's complicated structure. In developing the SDSFIE, multiple GIS communities are being served. Terminology used by one community does not necessarily translate to another community. The SDSFIE requires users to learn new terminology in place of what is normally used in their area of expertise.

Solutions

The CADD/GIS Center should focus efforts on technology transfer and implementation of their standards. Because of the breadth of the standard, many fields are not necessary for each discipline. Training information or subsets of the standards would help users. These could be extensions based on a user's mission area and type of duties; for example, hydrologists would be able to use the elements relative to their field and ignore irrelevant fields. This would work best in conjunction with alias tables that link what, for example, a hydrologist calls an item to what the standards call it. This would also allow specialists in many fields to use terminology familiar within their discipline, while still complying with the standard via a look-up table.

Problem: SDSFIE and GIS

Description

Another common difficulty for GIS users who try to use the SDSFIE is that it originated in an Intergraph environment, so ESRI GIS users find that the standard is inadequate to support their type of GIS data.

Solution

Future migration of the standard should take into account the ESRI data model to enable ESRI users an easy solution when implementing the SDSFIE.

Problem: SDSFIE and the Federal Geographic Data Committee standards

Description

Data producers in USACE are being told to organize their data dictionaries using the SDSFIE. They are under the impression that this conflicts with FGDC standards. Data producers feel that using both is redundant and time consuming.

Solution

Since the SDSFIE implements FGDS standards, there is no conflict. This needs to be explained to the field.

Problem: Refinement of SDSFIE

Description

It is thought by some data creators that creation of spatial databases adhering to the SDSFIE results in unwieldy, slow applications because of the required data structure of the SDSFIE.

Solution

A high-level panel on the SDSFIE should be formed to determine if revision would be useful.

Problem: File-naming conventions

Description

File-naming conventions are important to a GIS implementation. Where there is no organization of file names or coordination of file names between organizational elements, available data may be overlooked or lost.

Solutions

Many sections have naming conventions related to their file names and have had them for years. These conventions are given to new employees on their first day, so the schema is not broken with personnel turnover.

The Regional Engineering and Environmental Geospatial Information System (REEGIS) utilizes Intergraph's MGE forced data directory structure. Each REEGIS category has a two-letter code. The hydrosurvey portion uses year-rivermile.dgn as the file name convention. Another implementation is using projection, county, and type of data as the directory tree structure so users gain information about the data without having to look at the actual data. Once the data layer is identified, the metadata file provides the appropriate completed information so that the user is able to use the information correctly and intelligently.

It is suggested that the year of the data file be incorporated into the file name if the data and metadata are stored separately. If the data is updated more frequently than yearly, then include the month, day, or even time of collection. The use of naming conventions for file names decreases the amount of time necessary to find data relevant to a project. Other USACE elements are encouraged to incorporate naming conventions along with data dictionary and metadata standard use into their GIS implementation for optimal results.

Problem: Geospatial data documentation

Description

Use of the Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata within USACE varies by command and by disciplines within the Districts, Divisions, and laboratories. Officially, most Districts state they use the FGDC metadata standard and use CORPSMET software created to support the creation of metadata by USACE personnel. In conversation with various USACE personnel, it became obvious that many USACE employees do not realize the benefit of the metadata standard and do not use it. Many Districts use their own "proprietary" schemes for metadata documentation. In most cases, the leap from these to full FGDC metadata files is a small but vital step for sharing data beyond the organization's walls.

USACE personnel see the metadata requirement as another unfunded mandate. What they don't realize is that documenting their work is part of being a USACE employee. Excuses for not doing metadata ranged from "you can't make me do it" to "the customer did not ask for it." In the same way that USACE employees must use CEFMS to manage a project's finances, the requirement to create geospatial metadata is not an extra duty—it is one of the steps in a geospatial data creation project.

Solutions

For a more positive attitude about geospatial metadata to take hold in USACE, a conscious effort must be made to educate the project managers so the process includes GIS data and its associated documentation. This should dilute the perception that metadata is not necessary.

Some Districts have been proactive in metadata creation and dissemination. In some instances, metadata files have saved tens of thousands of dollars because they provided the information necessary for a vital project. In one instance, a contractor knew more about the metadata standard than the COTR did and was happy to provide the information since they saw it as a vital portion of the project. Some Districts have a vast collection of data and metadata on the Web; since the data are available electron-

ically, the public can download the data without involving coordination with a USACE employee. This decreases the number of public calls employees must answer, so they can devote more time to their other responsibilities.

Geospatial data library

A data library is an organized information structure for storing and archiving a collection of data. This facilitates the retrieval of data when it is needed. Associating data with appropriate keywords, and ideally with geospatial location, allows discovery of the data by others besides those who created the data. The investment in data creation is enhanced immeasurably when the data is stored within a data library.

Description

Data libraries are not prevalent in the USACE. Some Districts have lists of available data, but not in a true library. At times, the list of data themes is available electronically; in other instances it is in the memory of one or two knowledgeable users. Those questioned about data libraries believe that creation of data libraries would be an excellent asset to Districts to facilitate determination of what data are available and what data are needed for a project. One District is actively pursuing the creation of a data library.

Solutions

Standard procedures related to data acquisition are a key step to a digital data library. These procedures may include contracting language requiring metadata from the contractor, use of a standard data dictionary, format, and accuracy specifications for each data layer. Currently, Engineer Regulations and Engineer Manuals are used for theme-specific guidance related to data acquisition, but these lack overall District procedures related to the data.

A key factor in the use of a data library is the associated data documentation or metadata. Standardized metadata creation should be part of the business process of data creation.

At one office, a data manager position was created to coordinate the GD&S effort, including the development of a data library. The data manager was funded partially out of this work unit to facilitate a test of the concept. Unfortunately, the person originally slated for the position left employment in USACE. Afterwards, local managers found it impossible to allow any of their employees to leave their current duties to perform the role of data manager. The funding was used to enable a de facto GD&S manager to spend time writing a summary of data management in the organization (Appendix E).

Data sharing

Data are created to be used. The more often and the longer they are used, the more return on the original investment. The more ways that they can be used, the more valuable the product. However, sometimes people who create the data find it hard to let go of it. The reasons are many: some fear the data will be misused, some never feel the dataset is complete, some have personal issues with other individuals and do not want the others to profit from their efforts. These human issues are addressed here along with the technological and procedural difficulties in data sharing.

Problem: Technical hurdles

Description

Sharing data within a District or Division requires compatible hardware, software, and network configurations. One of the major obstacles is file format consideration. CADD is usually housed in Engineering, whereas GIS is normally in Planning or Information Management. Because the systems grew from different perspectives, different hardware and software systems of choice are used. Most CADD users favor Autodesk's AutoCAD or Bentley's Microstation software. GIS users normally favor ESRI's ArcInfo or Intergraph's MGE software packages and their associated viewing tools. New GIS and CADD packages have recently come on the market, but they have not seen widespread use in the USACE.

The CADD and GIS software industry have made it easier to use data from other native dataset environments by including data readers or extensions that read other data formats. Although the problem of viewing cross-application or cross-platform data seems to have been solved, the way the different environments treat the data still varies. For example, the treatments of annotations and line widths differ in CADD and GIS, and this type of information is often lost when viewing the data outside of the native dataset environment.

Solutions

Software vendors are addressing this issue. They have developed software used in conjunction with relational database management systems (RDBMS) to allow data to be served to a variety of software and hardware platforms. The two most prevalent software products are ESRI's Spatial Database Engine (SDE) and Bentley's Continuum. These products allow CADD users to see the data in CADD format while GIS users see the same data in GIS format without any loss of information; they allow widespread data sharing while maintaining the integrity of the

data by using RDBMS permissions to control access and updates. Multiple versions of the data are also eliminated when using this type of software. Another benefit is the maintenance of skills, macros, and other customized tools acquired during the life cycle of GIS and CADD use at the District. The Geospatial Data and Systems work unit tested the usability of the SDE product within the Mississippi Valley Division. The results of this demonstration are available as a supplement to this report (Appendix D).

Problem: Cultural hurdles

Description

Technical issues related to sharing of data across an organization are within a few development years of being resolved, but the human and political issues related to true data sharing will take more effort. At some surveyed Districts, data sharing is based more on who you know than on another paradigm. There are territorial disputes between various USACE sections to the detriment of the organization.

Solutions

Sharing of data requires a certain level of trust, for example, trust that the data will be used appropriately, trust that an in-kind sharing will be forthcoming, and trust that each group will be included in future data procurement. This trust should be fostered by the managers as well as the individuals involved. Cooperative work via the Technical Committees should help alleviate this problem. Data procurement cost sharing allows groups to get better or larger datasets because of cost sharing and leveraging. These datasets are of greater benefit to many sections in a District. A corporate approach to data acquisition should be used whenever possible.

Problem: Version control

Description

Version control is a primary concern to USACE as the number of data users for both CADD and GIS data increases. Geospatial data are often saved on local or personal computers as changes are made to a data layer. These changes may be made as another user simultaneously makes changes to the same data layer that she has downloaded to her personal computer. Since two copies of the data layer are being edited at the same time, it is difficult to discern which is the master data layer. When the error is discovered, one person's work may have to be duplicated.

Solutions

In the CADD arena, the Tri-Service CADD/GIS Center has developed guidelines for the implementation of an electronic document management system (EDMS). The guidance may be downloaded from http://tsc.wes.army.mil. The Corps of Engineers EDMS has been successfully deployed at the Jacksonville District, and it manages more than 100 Microstation licenses. The system has been incorporated into the workflow and maintains a master copy of drawings.

GIS implementations of geospatial data management systems have not been as prevalent. The GIS realm is a relatively new market, and most products are still available only in version 1.X. These systems, such as ESRI's Spatial Database Engine (SDE) and Bentley's Continuum, use a centralized server with relational database technology. They provide control, access, update authority, and attribute domain checking as well as the data in a variety of formats. Some are beginning to allow for multiple copies of the same data layer. They maintain the integrity of the data due to the security of the data and limit change authorities to thematic functional area experts. As part of this work uni, a test was conducted using SDE within the Mississippi Valley Division. The write-up of this demonstration project is available as a supplement to this report (Appendix D).

Optimizing geospatial data work flow and document flow

There has been some discussion on requiring the use of pre-existing document storage systems that were not created specifically for use with GD&S to "optimize" document and work flow in the USACE GD&S arena. It certainly would be valuable to improve work flow, for example by making metadata creation a standard and funded part of the GD&S workflow. It would be valuable to improve document flow by channeling all completed data collections into geospatial data libraries. However, we believe that the use of pre-existing work flow or document storage systems that are not designed for use with GD&S will inhibit the productivity of USACE GD&S employees.

The description below shows that the diversity of GD&S work, document, and data flow in USACE belies the possiblity that one solution will fit all jobs. The diversity is due to the nature of the work being done. It is the responsibility of USACE employees involved in each GD&S project to optimize the flow of that project.

Problem: Work, document, and data flow

Description

Requests for GD&S products come from the military, other federal, state, local, commercial agencies, other USACE elements, or within the District or Divi-

sion. It may enter the system via a telephone, hard copy, email, or person-to-person contact, or it may be part of a long-term understanding that the product will be produced. It may be work to be done by USACE employees or to be contracted out.

Resource data used to produce GD&S products comes from field surveys, remote field data collection platforms, remote-sensing platforms that are in place or that need to be scheduled, or currently available GD&S products including map or database materials from federal, state, local, or commercial groups. Data may come in as raw data products in electronic files, massaged data in electronic files, maps in electronic files or hard copy, or textual information in electronic files or on hard copy, all in various file formats.

GD&S products created in USACE include maps in electronic or hard-copy format (produced from electronic format files). An electronic map often consists of multiple data files. Sometimes corresponding metadata is produced with map data. While Federal Executive Order 12906 and USACE ER 1110-1-8156 require that metadata must be created for any new geospatial datasets created by USACE organizations, metadata may or may not be created, depending on the organization, individual knowledge of metadata, funding, and workload. Textual reports may accompany GD&S products in electronic or hard-copy format.

GD&S products may be distributed directly to the requester, made available to the public, or both. Products that are paid for by a particular funding agency are supplied only to that agency. Responsibilities for further distribution lie with the owning agency. Often further refinements or updates to the original products are requested—the work is not necessarily over after the first product is complete.

GD&S electronic map products are usually large electronic files, up to hundreds of megabytes in size. In some cases, storage of a collection of related products requires terabytes of computer memory. (This does not include the working drafts required to get to a finished product.)

Products may be archived off-line, stored on-line for real-time or near real-time downloading by in-house personnel or the public, or provided to the public using non-electronic media (CDs, various tape formats, or hard copy). Geospatial data can be stored and made public on the USACE Geospatial Data Server, which is accessible to USACE and the public via the internet (http://corpsgeo1.usace.army.mil), and has been created specifically for this purpose. The Server will have access to the CEAP tape farm when storage needs require that interface.

Storage schedules, platforms, and media for USACE geospatial data and metadata are organized and implemented by the USACE elements creating the products. GD&S products paid for by non-USACE agencies are the responsibility of those agencies. During site visits, we found that USACE GD&S data producers are aware that they have a responsibility to make data that is created with public funding available to the public. Indeed, this is widely seen as a good public relations tool and is looked on favorably by both managers and staff.

Solutions

Work and document flow in the various USACE GD&S work environments varies according to the reason for the GD&S work, data collection schedules, products required, customers, offices involved, and available funding sources. There is no one formula that will work for every scenario if we are to keep the GD&S work environments optimally efficient. A document management system that is not specifically designed for geospatial data flow could not be appropriate.

Vice President Al Gore wrote about the U.S. government (Gore, 1995),

> ...we have created a system that demands that one size fit all, and in the pursuit of certainty we have created a system that attempts to cover every eventuality, spelling everything out in excruciating detail. ... In the words of Philip Howard, we have "exiled human judgment." But the world is neither all one size nor all that certain. Things change constantly, conditions vary, and human judgment is crucial to making things work. Or, as Howard puts it, "Decision making must be transferred from words on a page to people on the spot."

Nowhere are change and varying conditions more apparent than in the area of geospatial data production, which depicts the changing world and uses constantly changing computer tools to do so. GD&S products have innumerable preliminary iterations, each of which can have huge storage requirements. GD&S employees, the "people on the spot," must continue to be aware of their responsibility to publish data needed by the public on-line and to store archival data. They must be given the time and resources to do so. It is the responsibility of data creators and their managers to archive final versions of the data and metadata in GD&S storage systems that adequately support file sizes and file formats specific

to GD&S. Archiving of historical data should occur whenever possible, using existing business process storage methods.

The presence of a GD&S manager would increase efficiency of both the work and document flow within USACE GD&S work environments. Since no one organizational element is a proponent for GD&S in USACE organizations, there is a strong need for an individual to organize GD&S at the District level. A GD&S manager at each site, working with responsive management and a supportive IM, would be in a position to optimize the workflow within each work environment. For example, a GD&S manager could be responsible for:

- Creating data request forms for in-house and public use
- Planning, with IM support, storage requirements, hardware/software/network update schedules, and other GD&S infrastructure planning for the group
- Ensuring the creation of responsible security and archiving plans for USACE GD&S data and metadata
- Creating specialized individual development plans (IDPs) for GD&S employees, to include data and metadata expertise
- Including appropriate USACE-required GD&S specifications in contracts for GD&S data and metadata production
- Ensuring the fulfillment of GD&S-related contract requirements
- Developing GD&S capabilities in a branch that is often making requests from another branch to produce GD&S products for them.

Responsibilities of a GD&S manager could be guided by Headquarters and determined by the Chiefs of the offices with GD&S requirements in each organization, based on input from their staff and/or input from the GD&S Technical and Oversight Committees. A GD&S manager would be in a position to organize each of the diverse GD&S environments in USACE, making best use of the sophisticated GD&S expertise of its employees, the established organizational structure, and USACE requirements to keep USACE GD&S on the cutting edge of service to the Army and other customers. Creating this position will necessitate setting aside resources to fund the position or the portion of a position dedicated to these tasks, since GD&S personnel are already stretched to the limit.

USACE geospatial data and systems management: Technical aspects

Data creation and manipulation

An amazing layering of protocols, hardware, systems software, and applications software must all function together smoothly for GD&S to work. In the following section, we summarize some of the planning, pitfalls, and success stories that we found during the survey.

Problem: Interoperability of hardware, software, and networking

Description

Interoperability between the hardware, operating systems, software, and network of systems of all GD&S participants is a requirement for GD&S success. There are no requirements within USACE at this time to use any particular hardware or operating system for geospatial data work. PCs running Windows 95, 98, or NT, workstations running some form of UNIX, and, to a lesser extent, Macintosh computers all exist side by side. This is a logical development, because different uses and users need different tools.

Solutions

Hardware interoperability between the computer systems of geospatial data creators and storage media is a problem that can be attacked by using appropriate software and networking. For the most part these can be networked using the appropriate software. For UNIX to Windows NT, 9x, or 2000 operating system interactions, the networking software SAMBA appeared to be the most successful; it allows PC users to interact with UNIX computers as if they were another drive on the PC. The software DAVE can be used to allow Macintosh folders to appear on NT PCs, and AppleTalk networking allows Macintosh users to view UNIX directories from their desktops and access them as if they were folders on their own desktops.

Problem: Interoperability of GIS data formats

Description

The interoperability of diverse data formats of diverse GIS and CADD systems are discussed in the administrative section of the project report. Users want to keep using the GIS or CADD systems in which they have already developed expertise. In some cases, this expertise may have been developed over many person-hours of USACE employment. It is logical and rational to maintain current systems rather than to completely replace them with new systems to force the entire Corps to run one GIS or CADD

system. In addition, we cannot contribute to a monopoly of any one geospatial software vendor to the total exclusion of others.

Solutions

To facilitate interoperability between geospatial data development systems, USACE should strongly support development of data translators, conformance to SDSFIE, and OpenGIS. It would be valuable to train GD&S personnel in these areas in PROSPECT GIS courses. GD&S users should have access to GIS-product viewers and image-viewers that are available as freeware or for purchase.

Problem: Information Management involvement in USACE GD&S

Description

In general, GD&S environments were less successful when IM personnel were not involved. Cases also exist where IM has not been involved and the technology environments were highly successful, but it is most efficient to have involved IM personnel working with GD&S staff to enable each to focus on their area of specialty.

Solutions

To foster this interaction, at least one IM representative should be an active member of the GD&S Technical Committee and should be responsible for the IM information section of the GD&S Implementation Plan and Plan updates. Ideally, the IM staff member would have experience using the geospatial software, hardware, network, and storage mechanisms of the organization. GD&S technology support should be part of his or her job description, and his or her job rating should rely in part on the IM staff member's support of the GD&S technology environment.

IM support of GD&S is also required at the management level. Presentations on the USACE GD&S IM needs and metadata requirements should be made to the Chiefs of Information Management (CIM) meetings on a continuing basis. A USACE Geospatial Data Advisory Group (GDAG) representative who is also in IM may be able to make these presentations to the CIM meetings.

Incorporating IM into the GD&S loop may be difficult in situations where IM has not been a part of the geospatial scenario thus far. In some cases IM has refused to support the hardware and software needs of the geospatial group because it didn't fit into IM's overall plan for computer systems infrastructure. This divisive attitude cannot be tolerated in today's competitive economy. IM managers should be held accountable for technological support of GD&S functions.

Problem: Technology and geospatial metadata creation

Description

Most problems related to metadata creation are not due to technology but to workload, funding, and resistance to using new computer tools. However, technology is involved: we need to make the technology of metadata creation function easily, quickly, and, as much as possible, invisibly to the users. Some people hesitate to use CORPSMET because it is another new program to learn and the metadata content standard terminology is obscure.

Solutions

IM involvement is crucial to the upkeep of networks, software, centralized data servers or repositories, and local computers. In addition, if training in computer-related classes were required of Corps staff, the boundaries of the "unknown" that many people fear would shrink. Computer training for GD&S staff should be an annual requirement to keep up with the latest information management tools and techniques, especially operating system and TCP/IP networking expertise.

All employees working in GD&S and all new employees in GD&S areas should be required to take a hands-on class in metadata creation. This class should be developed as part of the USACE PROSPECT training. The course requires more time than can be allowed for the subject matter in the PROSPECT GIS courses now given. It should also be given as a preliminary class prior to GD&S-related symposiums that Corps personnel attend. The only way to make metadata a part of the business process in USACE GD&S is to make it a part of the "normal" corporate thought process of geospatial data creation.

As is true with the concept of simultaneous data and metadata creation, if these tools are part of current college-level curricula, these issues may be resolved as people with less psychological resistance to metadata as an "added" work task come in. USACE should support teaching metadata concepts to all personnel coming up through the ranks of the educational and USACE systems.

Problem: Hardware limitations of CORPSMET

Description

CORPSMET is a public-domain software package owned and maintained by USACE. Its purpose is to provide a user-friendly interface for metadata cre-

ation. Unfortunately, it only exists for use on Windows-based PCs. This makes it less likely that the metadata will get created, because a large number of GD&S personnel are working on UNIX workstations due to the large processing capacity needed for GD&S work.

Solutions

For those with UNIX workstations or Macs, this problem can only be overcome by going to another computer to create metadata. CORPSMET will not be written for UNIX or Macintosh platforms in the foreseeable future. If there are non-PC users doing GD&S on site, they should have access to a public PC with CORPSMET installed on it.

The next version of CORPSMET should be written in a software language that makes it transportable to all platforms, or it should run as a database application with a Web interface that is available from all platforms. If CORPSMET is not going to be developed further, Headquarters should choose and standardize the Corps on another metadata creation platform that can be used on multiple systems.

The inclusion in GIS software of metadata creation modules is a valid alternative to CORPSMET if they produce metadata in the required Content Standard format described by the Federal Geospatial Data Committee. Metadata creation is then integrated into the normal workflow of data creation.

Problem: Putting metadata or data on USACE Geospatial Server

Description

We found that many USACE employees do not know how to create metadata or how to submit it to the USACE Geospatial Clearinghouse Node, which is aligned with the National Geospatial Data Clearinghouse. Metadata that is available on the USACE Server is simultaneously available through the National Geospatial Data Clearinghouse to anyone with an Internet connection. This supports Executive Order 12909 by offering the general public one place to go for federally owned geospatial data. If GD&S proponents don't put their metadata on the Server, people will not go to the National Clearinghouse and search for their data requirements. People will not find the data they need. This negative feedback loop reduces the chances of successfully offering centralized availability of geospatial metadata and data in the future.

Solution

Metadata is moved to the USACE Server using file transfer protocol (FTP) software. Some people are unfamiliar with the FTP commands and hesitate to use them. Training in computer-related classes for all GD&S staff would ease a situation that is caused only by unfamiliarity. Again, all employees and all new employees in GD&S areas should be required to take a specific class in metadata that covers metadata creation, submission to the USACE Server, and searching on the National Clearinghouse. Research should be done on simplifying USACE Server metadata submissions.

Data storage media

Data storage can be broken down into four main categories: active, online storage for immediate retrieval; near-real-time storage, where the request must be manually serviced prior to access of the data; archival storage; and backup. Each type of storage has different requirements and can be fulfilled by different hardware. Each must have its place in a properly executed data management system.

Problem: Online storage

Description

Active online storage is used when actively manipulating the data.

Solutions

This type of storage must be fast and have enough room to store the currently active datasets as well as one or more working copies of each. It is used by programs that manipulate, massage, or analyze data, generally with a person actively working with the system. For this reason, it must have fast access times with no excessive delays prior to retrieval. It must be capable of working with data discovery tools, as discussed above in the first problem, Data Discovery, under Geospatial Data and Systems Management. Currently, this need is being served by hard drives.

Problem: Near-real-time storage

Description

Near-real-time storage is used for commonly accessed datasets. These are datasets that do not change frequently but are referenced, copied, or used by users and other datasets often.

Solutions

Near-real-time storage must be readily available, but immediate. Real-time access is not necessary. It does not need to be as fast as real-time storage, but it must be reliable. It must also work with the data discovery tools. Currently two main types of hardware are serving these data needs: hard drives and disk

arrays, and CD-ROM juke boxes.

In this arena hard drives and disk arrays have the advantage of flexibility and speed, which is not necessarily as important as reliability. Most hard drives fail within 3 to 5 years. CD-ROMs have the advantage of reliability with a sacrifice to speed and flexibility. If a dataset must be changed for any reason, the CD must be recut. Because CDs are read-only, they do not need to be backed up, although if you are cutting a CD you generally cut more than one, which serves as a backup itself. Properly executed, either is a viable option. In current technology, a large disk array properly backed up is generally easier to set up, administer, and use. It provides the speed users want with the flexibility required by administrators. Large hard drives are currently much cheaper than implementing large CD-ROM collections.

Problem: Archival storage

Description

Archival storage methods are used for datasets that are no longer active. Frequently these are datasets that have been superseded or outdated for some reason.

Solutions

These datasets should be kept for historical reasons and to fulfill community requests; instant access is not necessary. Archival storage is generally met by tape media in one form or another or by CD-ROMs. CD-ROM is the better choice for this type of storage in most cases, although it is initially more expensive.

Although this type of storage is easy to fulfill with tape media, there is a frequently unnoticed problem associated with it. Due to the long period of time between requests for this data, the media on which it is stored can easily exceed its intended lifetime, or become outdated, with no means to read it back onto a useful medium for transmittal. Several of the sites visited had wall racks of tapes that could only be read on one 10–15-year-old computer. When that computer fails, the information stored on the tapes will likely become inaccessible.

Newer media like 8 mm, DAT, and DTL tapes are only rated for 3 to 5 years. After this time, the media can fail or lose data without warning. CD-ROMS do not currently have a known life span. Theoretically, they can retain their data indefinitely, and have thus far proved reliable over periods of greater than 10 years. CD-ROMs have the added advantage of being considerably faster for retrieval of data than most forms of tape archiving.

If tape storage is decided upon, then the data on the tapes should be regularly validated and the storage media upgraded as needed. Move the data from obsolete media to current standards.

Currently write/serve jukeboxes exist for both CD-ROM and DVD. This eliminates dependence on external readers and allows batch conversion and renewal.

Problem: Backup storage

Description

Backup storage is used to recover from accidental data loss, unauthorized tampering, and disasters. Data are important. Data are money. Loss of data is a loss of productivity and money. Backup of data is the cornerstone of any good data storage methodology. Any decisions made regarding the types of data storage described above must take into consideration the eventual need to back up and restore data. All sites visited that performed regular backups used some form of tape media.

Solutions

Tapes are reliable (within their accepted limitations), cheap, and easy to use and administer. The choice of media is a quickly changing target. Most sites used a variety of 8-mm helical scan tapes that have been an industry standard for several years. This medium is typically good for storage of up to 14 Gb of data.

The newer standard is Digital Linear Tape (DLT). This medium is more reliable, has a larger capacity, and is considerably faster given the correct circumstances. If no standard currently exists at your site, then DTL should be the medium implemented for backups.

One important step in a successful backup strategy that most sites failed to take was validation of backups. Tape media fails, especially if used/rewritten often. Tape writers fail. Networks fail. Backups made regularly should occasionally be checked for validity. Can you restore data from the tape, and are they in the same condition as when they were written? Tapes should be retired before they reach their useful end of life so that loss of data will not occur. These steps will help to ensure that the data collected today will be available when you want it tomorrow.

Network architecture

Underlying data storage concerns are concerns about the network architecture deployed at the site. Many of the decisions to be made are based on speed of delivery. The type of active online storage chosen, the medium for backups, and archiving are all tied to the network upon which the data will flow. In addition, the choice of whether to keep files on individuals' local hard drives or to store them centrally relies

heavily on the underlying architecture.

Problem: Network operating speed

Description

It was found that 10-Mb networks are no longer fast enough to serve the needs of most medium- to large-sized offices. DLT drives do not operate efficiently at 10-Mb network speeds; file transfer speed hampers access to data and backups.

Solutions

Sites should implement a 100-Mb network as soon as feasible. Management at all levels should realize that if we don't upgrade our network capabilities, our geospatial capabilities will be severely handicapped.

It has been found, however, that once workstation segments are upgraded to 100 BaseT, the 100 BaseT segment to the GD&S Server becomes overloaded. Three-tier networking can overcome this problem: 100 BaseT for clients, Gigabit or ATM for servers, and a storage area network for backups. Two diagrams showing sample network architectures are available in Appendix G.

Problem: Centralized vs. distributed storage and manipulation of data

Description

Problems can arise when two people have copies of the same dataset on their own PCs and make changes in the data. The data may have been downloaded from a central data storage server across the network. If the data are not returned to the central server, each person loses the benefit of having the other's updates. If they both return the data to a central server, one may overwrite the other's changes or there may be two versions of the same data on the central server. There may not be a way to combine the changes made on each version of the data. The changes made in each of the data files may not even be known to the other person working on the same data.

Solutions

Three common scenarios were found for data storage and manipulation:

- Central storage and manipulation
- Central storage and local download, manipulation, and upload
- Central storage and local download with no replacement of data to the central server.

Centralized storage schemes rely on a small number of network hosts to store the data. Centralized storage enables easier implementation of data discovery software. Backups are generally faster and more stable. Recovery from disasters and hardware failure is faster and easier. It allows for a buildup of storage devices in one location where they can all be monitored easily and efficiently.

With 100-Mb network architecture, it was found that the options for data storage and manipulation are considerably more flexible. With a properly configured 100-Mb network, access to data on remote resources for manipulation appeared to be nearly as fast as access to locally stored information.

Local workstation computers may connect to the network hosts to retrieve and work with data. On some sites, data managers have instituted software to facilitate a centralized data repository with controlled access. Data is stored in the central database and can be "checked out" to only one user at a time. Thus coordination is enforced, eliminating creation of multiple versions of one data file.

At some sites data creators and manipulators prefer to keep the data on their local PC while they are working on it rather than in a shared database that they would have to access across a network. This may be due to slow networking, unfamiliar downloading interfaces, or just a personal preference to have possession of the data file while working on it.

Our study showed that in some cases version control through informal communications between those working on the same data was very well coordinated. This works well for organizations with a high degree of communications between GD&S users. As an organization becomes larger or as GIS becomes more distributed between offices and branches, the communication system is likely to break down. Long-term storage of local copies of data should be avoided. Upgraded network communications might make it more likely that users would work on centralized datasets.

Problem: Network protocols

Description

There are a number of network protocols that can be used over the physical network.

Solutions

The choice of protocols to lay on top of this physical layer should be limited to two main options: Microsoft networking and TCP/IP. TCP/IP is the industry standard for interoperable networking. As such it works well with UNIX systems, desktop and server PCs, and Macintosh computers, as well as older legacy computers. Microsoft networking has a narrower audience, being confined mostly to Windows PCs. However, many sites reported great success with a product called SAMBA, which allows UNIX-based servers to communicate with Windows-based PCs

using Windows networking directly. These options should be viewed as the only ones currently viable, except to access legacy data, until such time as the data can be transferred to a more current, standard system. Network File System was looked at by some sites but was found to be slower and less reliable than the others.

System security

Problem: System security standard operating procedures

Description

The most serious detriment at all sites was a lack of standard system security plans. Most sites had no standard operating procedures (SOPs) in place in case of disaster, hacker penetration, or physical computer crash.

Solutions

Due to the variety of available operating systems, threat potentials, and work environments, the definition of SOPs is beyond the scope of this document, but there are several issues you should take into account when designing your SOPs. Data backups as mentioned in the Storage section should be included in any formal system security plan. All systems on which data are stored or processed should have an official security accreditation per Army Regulations.

Problem: Passwords and network security

Description

The lack of robust passwords on multi-user computers at most sites would make penetration by determined hackers relatively easy.

Solutions

Creation and distribution of passwords should be dealt with at the District or Division level by experienced system administrators. According to current standards, passwords should contain a combination of numbers and upper- and lowercase letters. Each Operating Activity should have a trained person who is responsible for monitoring the health of computer systems as well as running intrusion detection software. The Information Management Directorate, working closely with the data managers, best performs this function.

The Corps of Engineers Automation Plan Network (CEAPNET) has instituted a network security architecture to help prevent intrusion by hackers and provide an added level of security. As with most security measures in the computer field, this architecture should not be viewed as the only precautions needed. Application of good security procedures and disaster recovery SOPs will prevent many potential problems later.

For more information regarding system security and SOPs consult the USACE Internet Center of Expertise's (ICE) Guidance on System Security http://www.usace.army.mil/ice/.

CONCLUSIONS

1. The GD&S paradigm for the future includes intensive Information Management participation and mutual cooperation. This includes management-level IM as well as IM staff.

2. Information systems may be diverse on a local level but should be connected using optimal high-speed networking and should be made interoperable using software solutions and through database standardization.

3. Computer training should be an annual requirement for all GD&S staff. IM staff involved in GD&S should also be required to take GD&S training annually also. This is necessary to keep USACE at the competitive edge needed in this environment where technology that is two years old is outdated.

4. Metadata training should be a requirement for all current and all new GD&S staff and. Familiarity with the technology, as well as the Content Standard concepts, will help make metadata a way of thinking for all of USACE.

5. Types of media for a given task vary greatly, but all must be reliably backed up to ensure continuity of operations.

6. 100-Mb networking is the minimum for reliable, fast transfer of files and data. This should be the standard implemented at all sites.

7. Information system security is an increasingly prominent part of today's network environment. All sites should have a security manager and SOPs for disaster and intrusion.

LITERATURE CITED

Federal Geographic Data Committee (1995) Development of a national geospatial data framework. FGDC, Washington, D.C.

Frew, J., et al. (1996) The Alexandria Digital Library testbed. *D-Lib Magazine*, July/August.

Holh, P., Ed. (1998) GIS data conversion, strategies, techniques and management. In *Finding the Forest for the Trees: The Challenge of Combining Diverse Environmental Data, Selected Case Studies.* National Research Council.

National Research Council (1997) The future of spatial data and society; summary of a workshop.
Sutton, R. (1997) Implementation Plan for Jacksonville Geospatial Data and Systems. GD&S Technical Committee, South Atlantic Division Jacksonville District, U.S. Army Corps of Engineers
U.S. Army Corps of Engineers (1997) North Atlantic Division Philadelphia District, Technical Information Management System Final Report.
U.S. Army Corps of Engineers (1998) Electronic Document Management Systems (EDMSs) Technical Guidance, Draft Report, August.
Varon, E. (1998) Building the digital library. *Federal Computer Week*, 24 August.

WEB RESOURCES

Alexandria Digital Library
http://www.alexandria.ucsb.edu/
Core software technology
http://www.coresw.com/CST/
Federal Geographic Data Committee
http://fgdc.er.usgs.gov/fgdc.html
Tech Mall
http://techmall.com/
Digital Libraries Initiative projects
http://www.cise.nsf.gov/iis/dli_home.html
Australasian ImageNet Pty. Ltd. (AiNet)
http://www.ainet.com.au/

APPENDIX A: USACE GEOSPATIAL DATA AND SYSTEMS MANAGEMENT: MANAGER'S CHECKLIST

USACE GEOSPATIAL DATA AND SYSTEMS MANAGEMENT: ADMINISTRATION AND PERSONNEL ISSUES

Geospatial Data Management

Data discovery

Finding the data necessary for a project.

- ☐ Web-mapping software applications available
- ☐ Geospatial data manager position
- ☐ Local area network (LAN) sufficient to handle the size of these data files

Time and money

GIS requires a commitment of both from management

- ☐ Plan and budget for hardware and software maintenance
- ☐ Plan and budget for data maintenance requirements
- ☐ Multiple projects and managers willing to provide needed resources
- ☐ Lines of communication set up between managers and project leaders

Understanding the power of GIS

Lack of understanding about what GIS can do for the organization

- ☐ Have copies of other Districts' GD&S or GIS implementation plans
- ☐ Executive office and IM community are in communications loop

Communication and coordination

Unclear who is responsible for GD&S administration

- ☐ Data, applications, and analysis shared at the District level between sections, branches and teams
- ☐ District GD&S technical committee meetings held (EM 1110-1-2909)
- ☐ Inclusion of all Divisions, branches, or offices that have some part in geospatial data production, storage, or use

Responsibility for GD&S

Who is responsible for GD&S data maintenance?

- ☐ Funding for G&DS includes managing data maintenance
- ☐ Multiple users of GD&S are all responsible for budgeting for data maintenance
- ☐ New users of GD&S are given the needed training in data maintenance via classes or being detailed to on-site training
- ☐ One individual is GD&S manager, responsible for implementation of GD&S in District (including ensuring proper use of the system, providing application assistance to new users, engaging management in realizing the power of GIS, and marketing the District's GIS capabilities to potential new customers)

Contracting geospatial data work

Contracting officer does not know how to interpret USACE language related to GD&S data and standards

- ☐ For every geospatial dataset created, the deliverable includes a corresponding geospatial metadata file that conforms to standards set by the Federal Geographic Data Committee
- ☐ Close relationship between contracting office (CO) and contracting officer's technical representative (COTR)

Geospatial data and systems maintenance

Budgeting for data maintenance

- ☐ Data maintenance is put into project budgets for mission-required data
- ☐ Labor for updates is charged to projects and is included in project budget

Data backups

A wide variety of routines exist for geospatial data backups

- ☐ Responsible planning for data storage and backups is in place
- ☐ Relationship between geospatial data users and IM people strong

Software and hardware updates

- ☐ Hardware, software, and data updates are incorporated into budgets
- ☐ There is a plan for GD&S hardware, software and data updates
- ☐ The update plan is communicated to all levels of GD&S employees and managers

Geospatial data standards

Data dictionaries

Members of GD&S community may not be using same GD&S language

- ☐ GD&S employees are knowledgeable about Tri-Service Spatial Data Standard
- ☐ Advanced TSSDS users contribute their opinions to TSSDS effort

File-naming conventions

Without filename conventions, available data may be lost

- ☐ File-naming conventions are in place in organization
- ☐ File-naming conventions are widely known and used across subgroups within organization
- ☐ Year of data creation is incorporated into file-naming conventions
- ☐ File-naming conventions, data dictionary standards and metadata standards are all adhered to within organization's GD&S community

Geospatial data documentation

Many USACE employees do not realize benefit of geospatial metadata standard and do not use it.

- ☐ Project managers are knowledgable about geospatial metadata standard
- ☐ Geospatial metadata is incorporated into normal business processes in GD&S work at the organization

Geospatial data library

Would allow GD&S personnel to determine what data are already available and what must be created or bought for a project

- ☐ Contracting language requiring metadata from contractors is standard in GD&S contracts
- ☐ Standard data dictionary, format, and accuracy specifications for each data layer are used for every GD&S project
- ☐ Metadata or data documentation is created for every new GD&S dataset (unless it belongs to another organization and they have requested that no metadata be made)
- ☐ A data manager is responsible for organization's geospatial data library

Data sharing

Technical hurdles

Including compatible hardware, software, network, and file format configurations

- ☐ CADD and GIS software used within the Division include data readers or extensions that read each other's data formats
- ☐ Relational database management system (RDBMS) software is used so data can be served to a variety of software and hardware platforms, e.g., ESRI's Spatial Database Engine and Bentley's Continuum. (See Appendix D for a description of this solution applied in the Mississippi Valley Division)

Cultural hurdles

Human and political issues related to data sharing

- ☐ GD&S oversight committee fosters communication and trust between managers of groups using GD&S in organization
- ☐ Effort is made by managers to foster trust between GD&S groups in organization
- ☐ Cooperation is fostered between GD&S workers via participation in Technical Committee
- ☐ Data procurement cost sharing is practiced in organization

Version control

Multiple versions of a data file can result when same dataset is unknowingly downloaded and worked on by two employees simultaneously

- ☐ Corps of Engineers Electronic Document Management System (CEEDMS) guidance has been downloaded from http://tsc.wes.army.mil
- ☐ Other data version control software is used
- ☐ Plan is in place for data version control

Optimizing geospatial data work flow and document flow

Work, document, and data flow

One specific corporate work, document, and data flow doesn't exist in USACE GD&S community due to wide variety of mission and product requirements

- ☐ There is a GD&S manager or group working to increase efficiency of both work and document flow within our GD&S environments
- ☐ Management is responsive to optimizing workflow within our GD&S environments
- ☐ GD&S technical and oversight committees work to optimize work flow in GD&S arena
- ☐ IM is supportive of optimizing workflow within our GD&S environments

USACE geospatial data and systems management: Technical aspects

Data creation and manipulation

Interoperability of hardware, software, and networking—Higher GD&S interoperability results in higher GD&S productivity

- ☐ Relational database management system (RDBMS) software is used to allow data to be served to a variety of software and hardware platforms, e.g., ESRI's Spatial Database Engine and Bentley's Continuum. (See Appendix D for a description of this solution applied in the Mississippi Valley Division.)
- ☐ For UNIX–NT PC interactions, networking software SAMBA is recommended.
- ☐ For Macintosh–NT interactions, software DAVE can be used to allow Macintosh folders to appear on NT PCs
- ☐ For Macintosh–UNIX networking, AppleTalk networking is recommended
- ☐ Effort is being made to network computer systems of geospatial data creators and storage media using appropriate software and hardware

Interoperability of GIS data formats

- ☐ CADD and GIS software used within the Division include data readers or extensions that read each other's data formats. CADD and GIS software used within the Division include data readers or extensions that read each other's data formats
- ☐ Relational database management system (RDBMS) software is used so data can be served to a variety of software and hardware platforms, e.g., ESRI's Spatial Database Engine and Bentley's Continuum. (See Appendix D for a description of this solution applied in the Mississippi Valley Division.)
- ☐ GD&S users are knowledgable about potential solutions to this problem, e.g., use of data translators, conformance to Tri-Service Spatial Data Standards (TSSDS), and OpenGIS products

Information Management involvement in USACE GD&S

GD&S environments are often less successful when IM personnel are not involved

- ☐ At least one IM representative is active member of GD&S Technical Committee
- ☐ IM representative is responsible for content of IM information section of GD&S Implementation Plan and Plan updates
- ☐ (Ideal:) IM staff member has experience using organization's geospatial software, hardware, network, and storage mechanisms
- ☐ IM support of GD&S is strong at management level
- ☐ IM managers are accountable for technological support of GD&S functions

Technology and geospatial metadata creation

Technology of metadata creation function must be easy for users

- ☐ Managers of GD&S employees see value of maintaining their workforce's technical knowledge in area of metadata creation
- ☐ IM keeps networks, software, centralized data servers or repositories, and local computers working smoothly

In addition to being knowledgable about geospatial data creation and manipulation, GD&S users are well-trained in computer-related areas, including

- ☐ Conformance to filename conventions
- ☐ Metadata creation
- ☐ Data transfer processes such as FTP (file transfer protocol)
- ☐ Archiving and data storage standard practices

Hardware limitations of CORPSMET

CORPSMET is only written for use on Windows-based PCs.

- ☐ Non-PC users doing GD&S have access to a public PC with CORPSMET installed

Putting metadata or data on USACE Geospatial Server

Many USACE employees don't know how to submit metadata to USACE Geospatial Clearinghouse Node

- ☐ GD&S users are trained to use FTP to move metadata (and data if desired) to the USACE Geospatial Data Server.

Data storage media

Online storage

Used when actively manipulating data. Organization's online storage

- ☐ Is fast
- ☐ Has enough room to store currently active datasets, as well as several working copies of each
- ☐ Works with data discovery tools

Near-real-time storage

Used for commonly accessed datasets. Near-real-time storage

- ☐ Is readily available
- ☐ Is reliable
- ☐ Works with data discovery tools

Archival storage

Used for datasets that are no longer active.

- ☐ Appropriate data are kept for historical reasons and to fulfill future requests for historical data
- ☐ For any media used for long-term storage, lifetime of media and duration of readability of media is considered in storage plan
- ☐ Storage *tapes* of any kind are regularly validated and medium of storage upgraded as needed.

Backup storage

Used to recover from accidental data loss, unauthorized tampering, and disaster recovery

- ☐ Backup storage media type has been carefully considered
- ☐ Backup processes are validated regularly by checking readability of data

Network architecture

Network operating speed

10-Mb networks are no longer fast enough to serve the needs of most medium- to large-sized offices

- ☐ 100-Mb network is in place
- ☐ There is at least a plan to put a 100-Mb network in place

Centralized vs. distributed storage and manipulation of data

See also Data Sharing: Version Control

- ☐ Centralized storage scheme is in place (EDMS software or other)
- ☐ 100-Mb network architecture is in place
- ☐ GD&S technical committee is highly sensitive to this issue
- ☐ GD&S community is highly sensitive to this issue

Network protocols

- ☐ Microsoft networking (Windows PCs) is in place OR
- ☐ TCP/IP (UNIX, desktop and server PCs, and Macintosh computers, as well as older legacy computers) is in place OR
- ☐ SAMBA (UNIX–Windows networking) is in place

System security

System security standard operating procedures

- ☐ Data backups are included in our formal system security plan.
- ☐ All systems on which data are stored or processed have official accreditation per Army Regulations.

Passwords and network security

- ☐ Creation and distribution of passwords is dealt with at District or Division level by experienced system administrators.
- ☐ A person trained in security is responsible for monitoring the health of computer systems and running intrusion detection software

- ☐ IM is involved in security of GD&S data holdings
- ☐ We are aware of Corps of Engineers Automation Plan Network (CEAPNet)'s plan for network security architecture

APPENDIX B: SAMPLE HARDWARE CONFIGURATIONS AND PROCEDURAL RECOMMENDATIONS

CHAD ADAMS

With a maximum life cycle of only about two to four years, today's computer technology moves at a very fast pace. Specific system recommendations are beyond the scope of this report, due to their ephemeral nature. Working closely with IM, the geospatial community should try to keep systems current to within 12–18 months of cutting-edge technology. Computers older than this will hinder performance and cost more money in lost person-hours and maintenance than buying a new computer. The minimal configuration should include a 100-Mb-network interface and sufficient disk space to store the largest project to be worked on.

Sample optimal configuration specifications

Sun File Server—Sun Ultra 10, including system disks as needed, 18- to 36-Gb hard drives (as needed for data storage), 1 Gb memory or more; Solaris 2.7, Netatalk, SAMBA; version control software (optional), 100-Mbit-network interface, and DLT 7000 Tape Jukebox

Sun Client—Sun Ultra 10, including system disks as needed, 18-Gb data storage hard drive (as needed), 512-Mb memory; 21-in. monitor, Solaris 2.7, required data manipulation software, and 100-Mbit-network interface.

PC Client—800-MHz PC, including Windows NT 4.0, 256-Mb memory, 18- or 36-Gb hard drive, 21-in. monitor; required data manipulation software, and 100-Mbit-network interface.

Mac Client—333-MHz G3, including 18-Gb disk, 300-Mb RAM, Mac OS 8.X, 21-in. monitor, required data manipulation software, and 100-Mbit-network interface.

Network—100 Base T

Sample standard operating procedures

Backup schedule—Full backup performed once weekly. Incremental backups performed nightly. One tape each month kept for archival purposes. Weekly tapes should be stored for at least a month. All tapes should be kept off-site or in a fireproof safe. This type of schedule allows complete restoration of all files to within 1 day with the restore of only two tapes while still keeping down the number of tapes required.

Security—The local security manager should be tasked with checking patch levels for all operating systems involved at the site. Systems should be maintained at the highest patch level possible. An SOP should be developed for whom to contact during nonbusiness hours in case of the following events:

Hacker break-in—System administrators; Security Office; CERT; Public Affairs; CEAPNET Hotline.

Hardware failure (minor)—System administrator/repair person.

Hardware failure (major)—System administrator/repair person; CEAPNET Hotline.

Loss of network connectivity—Network administrator; CEAPNET Hotline.

APPENDIX C: GEOSPATIAL DATA MANAGEMENT SURVEY

The survey was conducted by research team members from CRREL and CERL on a one-on-one basis with representative Districts and Divisions. Five USACE commands participated in the survey: Jacksonville District, Mississippi Valley Division, Rock Island District, Walla Walla District, and Philadelphia District. Anonymity was promised to ensure open discussion of the issues.

The survey was conducted by at least two individuals at any one time. During the visits to each site, interviews were conducted between three to eight persons responsible for some facet of GD&S. No one functional element is responsible for GD&S at each organization, which stems from lack of USACE GD&S standard business practices. Interviewees were from the engineering, environmental planning, hydrology, information management, and real estate elements of their organizations.

Geospatial data management survey questions

Data management

1. What is the biggest data management issue at your District?
2. What is the most important type of data at your District?

Data maintenance

3. Does your District do routine data maintenance? or is it other duties as assigned?
4. How often do you feel your data is updated?
5. What data layer/theme gets the most attention as far as maintenance goes?
6. What data layer/theme gets the least attention related to maintenance?
7. Do you feel data maintenance guidance or standards related to theme and frequency of update would be followed at your District?
8. Do you 'throw away' data if an updated layer comes in?

Data backups/storage

9. What is frequency of data backups at your District for CADD/GIS/RS data?
10. Do certain themes warrant more frequent backups?

Equipment currentness

11. Does your District have difficulty using various versions of the same software packages?

Naming conventions

12. Does your District use any naming conventions related to file or directory naming?
13. Do you maintain the same theme of data collected at various times?
14. Do you incorporate the date of the data with the name of the file?
15. Is there a record of what is in each file? (metadata-light)
16. Do multiple people have a 'copy' of the same data layer?
17. Do multiple copies cause concern about what data is the most up to date? (version control)

Data library

18. Does your District have a catalog of its geospatial data?
19. Does your District utilize the metadata standard?
20. What is the biggest road block to full utilization of the metadata standard?
21. Is there an SOP related to data acquisition?
22. Does your District prefer to store internet-accessible data on a "corporate" server or a local server?

Data sharing

23. How do District personnel find useful data from within the District for their projects?
24. Is there a coordinated data acquisition group?
25. Is there any animosity about data sharing at your District related to the ownership of the data or fear of misuse?
26. Is file format an obstacle to the sharing of data between branches at your District?
27. What branch/division has primary responsibility for GIS?
28. What branch/division has primary responsibility for CADD?
29. Do territorial disputes hinder potential data-sharing endeavors at your District?
30. Is accuracy of data a big issue at your District?

Responsibility

31. Is there a person or group who is responsible for data, either tasked or ad hoc responsibility?

Data dictionary

32. Does your District utilize data dictionaries for ease in data sharing?
33. Has your District implemented any of the parts of the Tri-Service Standards?
34. What is the biggest benefit of the Tri-Service Standards to your District?
35. What is the biggest roadblock in full implementation of the Tri-Service Standards?

Active online storage

36. What type of physical media do you use for online storage?
    - 36.1. Single tape
    - 36.2. Robotic tape "farm"
    - 36.3. Hard disk
    - 36.4. CD (CDR)
    - 36.5. CD jukebox
    - 36.6. Other: _____
37. Do you have an upgrade path for this media in mind?
    - 37.1. If so, what are those plans? _____
38. What are your storage space needs?
    - 38.1. Current: _____
    - 38.2. Future incremental (monthly, yearly…) _____
39. Is there an expansion plan in place or is expansion handled on a crisis basis?
40. How far back do you keep historical data online?
    - 40.1. Is there a cutoff date?
    - 40.2. Do you only keep as much as you can store online?
    - 40.3. Do you back up historical data before deletion?
    - 40.4. Do you keep data forever?
41. What are your typical file sizes?
    - 41.1. Current: _____
    - 41.2. Future: _____
42. How quickly do you expect to be able to retrieve online data?
    - 42.1. Immediately upon discovery of need
    - 42.2. Within a few hours
    - 42.3. Within a day
43. What problems do you have with the current active storage arrangements?
44. What successes have you had?

Backup storage

45. What type of physical media do you use for backup storage?
    - 45.1. Single tape
    - 45.2. Robotic tape "farm"
    - 45.3. Hard disk
    - 45.4. CD (CDR)
    - 45.5. CD jukebox
    - 45.6. Other: _____
46. Do you have an upgrade path for this media in mind?
    - 46.1. If so, what are those plans? _____
47. What are your storage space needs?
    - 47.1. Current: _____
    - 47.2. Future incremental (monthly, yearly…) _____
48. Is there an expansion plan in place or is expansion handled on a crisis basis?
49. How far back do you keep historical data offline?
    - 49.1. Is there a cutoff date?
    - 49.2. Do you only keep as much as you can store online?
    - 49.3. Do you back up historical data before deletion?
    - 49.4. Do you keep data forever?
50. How quickly do you expect to be able to retrieve backed-up data?
    - 50.1. Immediately upon discovery of need
    - 50.2. Within a few hours
    - 50.3. Within a day
51. What problems do you have with the current backup storage arrangements?
52. What successes have you had?

Platform

53. On what type of platform do you store your online data?

	Data creation	Data storage	Data serving
Sun			
SGI			
Windows (3.1, 3.11)			
Windows 95/NT			
Macintosh			
Intergraph			
Other (specify)			

54. Do you have an upgrade path planned for these platforms?
    - 54.1. If so, will you upgrade to the same type of hardware or move to another type?
    - 54.2. What type: _____
55. What problems do you see with your current hardware platform?
56. What are its strengths?
57. Cross-platform interoperability:
    - 57.1. What types of systems can easily access your data?
    - 57.2. What limits easy access to this data?

57.3. What interoperability issues do you have?

58. Online storage costs:

	What was actual cost?	In hindsight, rank costs in order of importance (1 is most important)
Hardware purchase		
Software purchase		
Software development		
Training		
Maintenance: personhours		
Maintenance: media		
Maintenance: hardware		

Application

59. How do you offer your data online?

	Within your organization	To your customers
Web		
NFS		
FTP		
Other (specify)		

60. Do you have plans to upgrade this mechanism in the future?
61. Do you have a maximum file size for downloadable data? If so, what is it?
62 Is there a standard acceptable download time for data files that your users download?
    62.1. What is the maximum time per download that you plan for?
63. Are files downloaded immediately when requested by a user?
    63.1. If files are downloaded at a later time, what is time frame for downloading?
64. Do you use a specific application for providing data?
65. How do you index/catalog your data?
66. Are you satisfied with the current methods of serving and storing data?
    66.1. What improvements would you make?
67. Describe your current public "best seller"
    67.1. Theme _____
    67.2. Size _____
    67.3. Frequency of access by the public _____

Network topology

68. Where is your online storage located?
    68.1. Onsite
    68.2. Off site at a Corps processing center
    68.3. Off site on Corpsgeo1
    68.4. If off site:
        68.4.1. Is this data readily accessible to you?

68.4.2. Do you have data access/cooperation issues?

69. What protocol do you use to serve the data?
    69.1. TCP/IP
    69.2 Netbeui (Windows networking)
    69.3. AppleTalk/Ethertalk (Macintosh networking)
    69.4. Other _____
70. What type of internet connection do you have (T1, 56K…)?
    70.1. How does this affect size/type of file you can provide online to your customers?
71. What type of local network do you have?
    71.1. How does this affect size/type of file you can provide to your organization?
72. What type of network do your customers need to properly access your data?
73. What about the network topology hinders your efforts to access and serve data?
74. What are the good points about your network topology?

Physical security

75. What types of electrical problems do you guard against ?
    75.1. Power loss
        75.1.1. UPS
        75.1.2. Generator
        75.1.3. Other _____
    75.2. Electrical storms/surges
        75.2.1. Filtering UPS
        75.2.2. Surge protectors
        75.2.3. Other
76. How do you protect data from physical computer faults (crashes)?
    76.1. Client computers (data creators/users)
    76.2. Servers (data/backup)
77. Backups
    77.1. What data do you back up on client computers?
        77.1.1. Whole computers
        77.1.2. Data only
        77.1.3. Data and applications
    77.2. What data do you back up on server computers?
        77.2.1. Whole computers
        77.2.2. Data only
        77.2.3. Data and applications
    77.3. Do you have a regular backup schedule?
        77.3.1. Servers
        77.3.2. Client computers
    77.4. Do you perform validation checks of backups?
78. Do you provide off-site storage of backups in case of catastrophic loss (natural or other disaster)?
79. Unauthorized access threats (hacking):
    79.1. What type of computer security program do you have in place?
        79.1.1. Qualified computer security professional
        79.1.2. Security programs
        79.1.3. Logging/monitoring of connections
        79.1.4. Logging/monitoring of file downloads
    79.2. Do you apply the latest security patches/updates?
        79.2.1. For your operating systems?
        79.2.2. For your applications?

79.3. Which of the following do you consider a threat to your data/systems?
79.3.1. Intrusion from within your organization
79.3.2. Intrusion from other Corps sites
79.3.3. Intrusion from outside the Corps (other DoD, internet…)
79.4. How are passwords handled on your servers?
79.4.1. UPASS
79.4.2. Local district/division policy
79.4.3. Local branch/organization policy
79.4.4. Local machine policy
79.4.5. For your servers
79.4.6. For your client systems

80. Does your organization have any crisis management SOPs in place?
80.1. Disk/computer crashes
80.2. Archive loss (destruction)
80.3. Data alteration or loss
80.4. Viruses

81. Are you satisfied with the security measures your organization has in place?
81.1. Do they hinder your ability to function smoothly?
81.2. Do they provide you with a good feeling about the security of your data and systems

Miscellaneous questions

82. How many and what types of data formats do you provide to users?
82.1. Local
82.2. External customers

83. Data creation:
83.1. Who creates the data?
83.2. What is their relationship to the people who store the data?
83.3. How is data transferred to online storage?

84. File size:
84.1. How large are your typical datasets?
84.2. Do you store multiple-file datasets in one archive or as separate files?
84.3. Do you provide off-line copies of datasets?
84.3.1. For all datasets?
84.3.2. Only for very large datasets?

85. Data retrieval:
85.1. Who has the right to access online storage?
85.1.1. Data creators
85.1.2. Data owners (if other than creator)
85.1.3. Data maintainers
85.1.4. Customer for whom the data was created
85.1.5. Any customer with a similar need

Organizational questions

86. Who in the organization plans the hardware/software/media/network architecture for your geospatial data storage/manipulation needs?
87. What is the title/position of the person(s) who requests increases in storage capacity?
88. Who makes the final decisions of what will be purchased to increase storage capacity and processing hardware?
89. What is the organizational relationship of these people/groups?
90. On a scale of 1 to 10 (1 being lowest), how well would you say this system works?

APPENDIX D: SPATIAL DATA ENGINE PROTOTYPE FOR THE MISSISSIPPI VALLEY DIVISION

CHRIS REWERTS, CERL

Introduction

The Mississippi Valley Division (MVD) and the Environmental Systems Research Institute (ESRI) undertook a cooperative experiment to use ESRI software to manage MVD data. This experiment was supported in part by the US Army Corps of Engineers Construction Engineering Research Laboratories (CERL) for the purpose of gaining understanding of the software tools and their application to real data and operational situations to address a need. This document reports on observations made during the 12 May 1998 demonstration of the software usage prototype presented by ESRI at the MVD.

Background

Mississippi Valley Division

The MVD comprises all the US Army Corps of Engineers districts that have in common the Mississippi River System. Previously, the river flowed through two Corps divisions: the Upper Mississippi Valley region (referred to here as UMV) included the St. Paul and Rock Island Districts of the former North Central Division, and the former Lower Mississippi Valley region (referred to here as LMV) consisted of the Memphis, Vicksburg, and New Orleans Districts. What makes the joining of UMV and LMV notable in this report is that the format of spatial data used by the two are different; more specifically, the software used to create and maintain the data for the two previous divisions are different. With a new mission to focus collectively on the entire Mississippi River System, one of the problems faced by the MVD is how to share and integrate disparate formats of data. ESRI's SDE (Spatial Database Engine) software is being investigated as a technology that may provide part of the solution.

Data

The data are primarily geospatial or geographical information system (GIS) data. These spatial data are also connected, linked, integrated, or associated with tabular data. The Upper Mississippi Districts are using ESRI's ARC/Info as their GIS, with tabular data stored as attribute tables in the GIS. The LMV Districts are using Intergraph Corporation's MGE (Modular GIS Environment) for their GIS and CADD (computer-aided design and drawing) needs, with tabular data stored in SQL*Server (a Microsoft relational database management system) tables that are linked to spatial features stored in MGE. These components are part of the Regional Engineering and Environmental Geospatial Information System (REEGIS), which consolidates all engineering and environmental data for the Lower Mississippi River System into a standardized geospatial database.

SDE software

SDE is middleware software that provides linkage between GIS and RDBMS. Essentially, it enables storage of geospatial data in the RDBMS, allowing the exploitation of two mature software technologies, GIS/CADD and RDBMS, in concert. In addition, SDE stores and retrieves the data in such a way that they can be accessed and used by a variety of GIS softwares from various hardware architectures. For example, in the case of the MVD demonstration, ARC/Info data loaded into the SDE database could be accessed using not only ARC/Info but MGE as well. Without SDE, ARC/Info data would have to undergo arduous conversion to MGE format for this to be possible. SDE CADD client software enables MGE data to be loaded into the SDE database and then accessed by ARC/Info.

This description of SDE and SDE CADD client software is extremely brief, since much more thorough information is available at the ESRI Web site at http://www.esri.com/software/sde/index.html.

Experiment

ESRI and MVD agreed to demonstrate the use of ESRI's SDE software using sample data from the Rock Island and Vicksburg Districts, representative of the former UMV and LMV Divisions, respectively. The implementation design was for demonstration purposes and was not meant as a fully implemented system. The demonstration did, however, bring to light some of the many decisions necessary if SDE is chosen as a solution to the problem.

Observations

Since SDE is middleware, it provides little to observe by way of demonstration. If it is functioning with data loaded properly, it is virtually invisible. Most traditional types of live demonstrations of SDE in action on a computer, at events such as conferences, workshops, or vendor exhibits, can only provide an abstract description of what the software is doing, followed by the display of data of a GIS using SDE as a server. What one ends up seeing is data in an ARC/Info, ArcView, MGE, or other GIS or CADD display. To understand SDE, it is helpful to see it applied to a real-world problem, such as the sharing of the disparate data types of the UMV and LMV Divisions. A further difficulty in understanding SDE thoroughly is that one needs to have an understanding of the technologies with which it operates, namely the particular GIS, CADD, and RDBMS softwares with which it is to be used. It makes sense that a technology that offers a solution to such complex problems will be somewhat complex to implement.

The most important part of applying SDE to a given situation is planning. It is important to analyze the data that are to be loaded into SDE, so that proper choices can be made concerning geodata issues (e.g., projection) and RDBMS issues (e.g., a proper database schema). At the same time, considerations must be made based on how the data are involved in the process flows of the organization and the logistics implied. By way of example, several of the decisions faced in the MVD demonstration are described below.

Projection of geospatial data

UMV data were projected in UTM (Universal Transverse Mercator) while LMV used Mississippi State Plane. Options in such a situation include: reprojecting data to a common projection before loading it into SDE, using client software to project data as it is being served from SDE (if this is possible with the client software), or using the SDE projection engine "service" to handle projection "on the fly" as it serves the data. Of course, other issues of projection can come into play. For instance, the UTM data were in zone 15. If all data were to be projected to zone 15, this would have created problems, since all the data being joined may not lie in the same UTM zone. By the same token, all the data are not within Mississippi State Plane West. Projecting all data to decimal degrees may solve this problem in terms of combining everything in SDE, but this solution may be unacceptable for other reasons, such as incompatibility with established business practices or exchange of data products among users and customers.

Spatial indexing grid

When loading data into the SDE system, up to three spatial indexing grids can be defined. Choosing the size of the grid depends on the type of data, such as the size, shape, and density of features.

Database schema

One database scheme uses SDE to communicate with the DBMS to create tables to hold the data and then uses SDE to load the data into the tables. SDE groups its data into three types of tables in the DBMS:

- "F" tables, where features are stored

- "Business" tables, which store the feature attributes or tabular data
- "S" tables, which provide spatial indexing – one for each "F" table.

Meanwhile, since the attribute data are in the RDBMS, they may be linked with database queries to other tables.

This implies that thought must be given to how the tables are designed. In the MVD demonstration, this was handled differently for the UMV and LMV data. In the case of UMV, the process was straightforward. ARC/Info coverages were converted to ArcView shapefiles. These were loaded into the database with SDE functions that essentially create database tables with columns configured automatically to match the ArcView shapefile table designs.

The LMV REEGIS data required use of the SDE CADD client software because the data being used in the demonstration were created using MGE. Since the DBMS being used was the same that is used by the MGE system, the tabular portion of the data that were linked to the CADD data did not need to be loaded; elements that contained the MSLINKS before storage in the SDE system would be returned with the same linkages. The CADD client uses an interesting mechanism for dealing with the differences between GIS and CADD representations of features. For instance, CADD is based more in geometry, so that a center and radius may represent a circle object. Thus, if SDE is to be able to serve CADD data to a GIS, there must be some reasonable way to make the translation. The solution used by the CADD client is to create two columns for a given object in the table holding geospatial data. One column holds the CADD object exactly as it is represented in the CADD, the other holds a GIS-styled representation of the object. Thus, if a GIS requests the feature, it gets a representation it understands, but if the CADD requests the object, it can get the original CADD object as it was created.

Data update

Some geospatial data, such as state boundaries, are static; others, such as riverbank lines, change over time. Changes can include moving lines or points, such as in the case of a meandering river; adding or deleting features, for instance, when a control structure is installed or removed; or changing attributes, such as changing a name field when a new owner buys a land parcel. The impression we have of SDE is that it does not support dynamically changing geospatial data. The easiest type of change to make would be to attribute data, e.g., the changing of a parcel owner's name, since such a change could be done with a simple SQL (structured query language) change via the DBMS.

At the demonstration of SDE at the MVD, there was little discussion about how data updates might be accommodated. The overt implication was that all data were static, but in reality this would be rare. There will always be updates to a geospatial database such as MVD's, so these needs should be included in the design of the system. Depending on the frequency of changes and the need to publish the changes to the system, one way to deal with updates is by making changes over time to the "source" data in their native format, then at prescribed time intervals the old data in the SDE system could be replaced with the updated source data using the same steps to load it, including any projecting or transforming that was required.

Recall how the data in ARC/Info coverage format were first converted to ArcView shapefile format before loading into SDE. It is likely that if updates were to be made, it would be done with ARC/Info (rather than ArcView), since it has the GIS capability to ensure the geospatial data integrity. This brings up the important point that, while SDE is designed to handle geospatial data, it does not provide specialized functionality, such as building or maintaining topology. The design of SDE expects the program that was used to create the data (MGE, ARC/Info , etc.) is the "expert." Therefore, before data are loaded into SDE, they should be properly created and "cleaned."

Data standards

At the time of the MVD SDE demonstration, a looming issue for REEGIS was the Tri-

Services Spatial Data Standards (TSSDS). REEGIS has quite literally set the standard for the geospatial representation of river system mapping in the Corps, but how these would be melded into the TSSDS were yet to be defined. This does, at least, raise the point that consideration must be given to data standards such as TSSDS, metadata, or other such structures used to document data. It may be possible that in the future SDE could include an engine that would map data to TSSDS in a way that various clients (e.g., MGE, ARC/Info) could understand.

Problems encountered

It is rare that processes such as installing software, manipulating data, getting various software packages to operate and communicate, etc., occur without problems. At the time that this report is being written, we know that problems did occur and believe it is important to document the details, because they represent lessons learned. At this time, however, we do not have a full assessment of the demonstration in this respect.

Summary

The Mississippi Valley Division's problem of combining disparate geospatial data types into one system presents a complex problem. The existing data were expensive to create. The districts have invested in equipment and expertise using the GIS/CADD software of choice and would find it difficult to change. SDE offers a solution where the districts can continue without wholesale conversion of their geospatial data format. There are, of course, other technologies and possible solutions that are beyond the scope of this report. Whichever solution is chosen, much planning is necessary to design a system that will serve the organization well. It is hoped that this report has captured some of the key considerations to be made based on the 12 May 1998 SDE prototype demonstration at the Mississippi Valley Division.

APPENDIX E: JACKSONVILLE DISTRICT: DATA ARCHIVING, DATA MANAGEMENT, AND CONTINUITY OF OPERATIONS IN THE GEOSPATIAL DATA ARENA

RORY SUTTON, CESAJ

Several issues have recently emerged with regard to various collections of digital data. One, which is currently being referred to as "data management," is being raised in the context of federally funded environmental data. The broad idea is that federal funds are scarce and subject to accountability in a political, as well as a financial, sense. Scientific and other data financed this way must be safeguarded from loss and should be readily available to the research community in general. This maximizes return on investment, minimizes redundancy, and facilitates accountability. The Jacksonville District's participation in Everglades restoration and Florida Bay has brought us into contact with the problems associated with finding, obtaining, and maintaining environmental data in an interagency context. Preliminary discussion of data management has occurred in the Florida Bay PMC, the South Florida Ecosystem Restoration Task Force (SFERTF) Information Management Council, and SERA. There are federal initiatives. The NRC has published a book on environmental data management, *Finding the Forest in the Trees* (Holh 1998). Federal funding guidelines requiring return-on-investment (ROI) calculations and data management plans for recipient grantees appear to be coming soon.

Another issue that is asserting itself is the archiving of digital data that the District is required to maintain as permanent records. Examples of this are permitting actions under Rivers and Harbors legislation and the Section 404 program, hydrographic and topographic surveys, real estate determinations, and plans and specifications for projects. As one would expect, we have established procedures for complying with these rules. These procedures work well for physical documents, but they are starting to break down under the increasing digitalization of all aspects of our business process. Within the year, we will be capable of design, bid award, construction, payment, and transfer of a project without resort to hard copy.

Achieving this state of automation is an explicit goal of the Corps of Engineers and the District. At that point, the only reason to go to paper or mylar would be to comply with the old hard-copy records management rules. Another pressure on existing archival methods is decreasing space and budgets. The map file room, where hard-copy surveys are stored, is under constant pressure to become smaller or nonexistent. The warehouse, which stores real estate map files, is under consideration for elimination. These cost-cutting moves are sometimes made without provision for replacing the mandated archiving functions that were part of their reason for being.

Occasionally, the fact that a particular business process has been automated, and that records are now digital, is advanced as a reason for eliminating physical storage, without making provision for archiving the digital records. The assumption, it seems, is that if data is on a computer, then it is "archived." This assumption is baseless. Backup of District servers is designed only to provide reasonable assurance of overall continuity of operations under relatively minor hardware, software, and human failures. It is not designed for, and doesn't work well for, permanent storage and retrieval of essential records. The physical media used for backups is extremely volatile compared with paper. Even so, the useful life of a magnetic tape far exceeds the lifetime of the hardware and software required to read it. The Harris minicomputers were retired in 1991. A contract with Harris Corp. in 1995 was required to retrieve data still faithfully recorded on their 9-track tapes. Within a year, not even that will be an option. In this case, the problem is an unusual (these days) word size and compression scheme. Intergraph supplies software with their modern systems to read tapes made on the old DEC VAX systems. The difficulty now is the lack of a 9-track tape drive and an old Intergraph workstation to connect it to. The VAX was retired in 1993. Five years is a full generation in automation today, but a mere blink of the eye in

the archive business.

Continuity of operations (COOP) is a Corps policy requiring the District to continue to pursue its mission despite unforeseen events, which may go so far as to render its office space and computers, including onsite backups, unusable. A near miss by hurricanes Bertha and Fran during the 1996 hurricane season brought this possibility to the fore. COOP depends on off-site storage of essential backup data, as well as the possibility of finding a compatible computing plant to employ the stored data. Not long ago (1990), this meant sending a set of full monthly or quarterly backup tapes to Savannah District and receiving the same from them. The networked, distributed, heterogeneous computing environment we have migrated to since then has complicated any potential COOP solution. The software we use now (COTS) cannot generally be restored from backup, but must be installed where it is used. A variety of servers—Unix, NT, Novell—must be supplied to provide for system support and applications, and these vary widely from District to District. A bewildering variety of networking paraphernalia is required to establish the LAN and WAN connectivity we can no longer live without. These are difficult problems affecting the computing plant portion of the COOP equation. Of interest here is what COOP has in common with data management and archiving: the requirement for a secure, structured, catalogued, and documented collection of the data needed to get on with business.

The following is a draft of a plan to implement such a collection applied to the online digital geospatial data holdings of the Jacksonville District.

Data management for GD&S: The goal

The goal of GD&S data management is to identify, collect, organize, document, preserve, and make accessible the District's spatial data.

1. In consultation with project managers and technical staff, the GD&S committee, and possibly others, ***identify*** the locally developed spatial data (and related technical, scientific, and engineering data, if any) central to the District's projects.
2. ***Collect*** that data, as much as possible, on one of the District's GD&S servers.
3. ***Organize*** the data to maximize their utility and availability to the project team primarily, but also to facilitate their discovery and use by others. Intuitive file names and directory structures are important. Consultation with CADD and GIS staff will be necessary to preserve (or migrate in an orderly fashion) existing applications.
4. Produce, or otherwise obtain, FGDC-compliant metadata to ***document*** the collected and organized datasets, and post this to the USACE node of the NSDI. Maintain an online catalog to facilitate local access to this data.
5. Make offline archival copies of these data and their documentation so they are ***preserved*** for disaster recovery or for duplication and use by others. Arrange with IMO for a suitable backup schedule for the online data and a rewriting schedule for the archived data.
6. Respond to requests for data from the public and other agencies. ***Make data*** and its documentation ***accessible*** via Web-indexed files on the public FTP site (where that is appropriate), by providing the data on tape or CD-ROM, or by referring the requestor to the proponent or OC, if necessary.

Data management for GD&S: The plan

The primary unit of data management will be the *volume.* A volume consists of related datasets, exclusively occupying a compact disk store, along with their documentation, symbol sets, saved views, extensions, macros, and other programs operating on them. All these objects will be contained in a single subtree of the operating system directory structure. Thus they can be archived to a minimal saveset and restored with minimal loss of functionality due to dependence on other savesets. Each volume will be indexed by a "catalog" consisting of a brief description of each dataset and containing a pointer to detailed metadata for that dataset.

Volumes will be archived initially as soon as they are identified and organized. After the initial push, archiving will occur quarterly. Each quarter, one fourth of the volumes identified for data management will be archived, and the archives will be sent out for copying and offsite storage. Thus, each volume will be archived once a year while it is online. The existing backup retention policy is one year, so worst-case recovery would be a series of restorals from archive and backups. Archive retention policy will be "until superseded." Archived savesets will be refreshed by restoring to online media, performing some sanity checks, and re-archiving to new media every two years. Hardware and software procurement and disposal cycles will be constrained to not "strand" archives on unmountable media or in unreadable formats.

Hardware required
2 EXB-210 8-mm tape libraries (currently used for backup)
1 HP DLT-7000 tape library (in procurement)
20 Gb of scratch disk space on a Unix file server

Software required
Legato networker with Jukebox module (currently used for backup)
Legato archive module
Automated cataloging script (being locally written, 75% complete)

Currently identified spatial data volumes in SAJ

Florida basemap and emergency management data
Path: marvin:/marvin3/mapping/district
Path: marvin:/marvin3/mapping/fl_utm17
Path: marvin:/marvin3/mapping/fl_utm16
Path: marvin:/marvin3/mapping/ga_utm17
Path: marvin:/marvin3/mapping/doqq_img
Path: marvin:/marvin3/mapping/quads_jpg
Size: 17 Gb
Tapes: 2
Cost:

Puerto Rico basemap and emergency management data
Path: marvin:/marvin3/mapping/pr_lib
Size: 10 Gb
Tapes: 1
Cost:

USVI orthophotography, basemap, and emergency management data
Path: marvin:/marvin3/mapping/vi_lib
Path: marvin:/marvin1/usvi
Size: 4 Gb
Tapes: 1
Cost:

Kissimmee River Restoration orthophotography, topography, and project features
Path: marvin:/marvin2/kiss
Size: 6 Gb
Tapes: 1
Cost:

Central and Southern Florida Project and Restudy

Path: marvin:/marvin2/csf_restudy
Size: 2 Gb
Tapes: 1
Cost:

Upper St. Johns River Environmental Restoration
Path:
Size:
Tapes:
Cost:

C-111 and Modified Water Deliveries to Everglades National Park
Path: marvin:/marvin2/c111
Size: 2 Gb
Tapes: 1
Cost:

Lake Okeechobee Emergency Action Plan
Path: marvin:/marvin2/okeechob
Size: 2 Gb
Tapes: 1
Cost:

Hurricane Tracking
Path: saj3k:/saj3k1
Size: 4 Gb
Tapes: 1
Cost:

APPENDIX F: RELATED WORK: THE ALEXANDRIA DIGITAL LIBRARY PROJECT AND ON-LINE ACCESS TO INFORMATION VIA GEOGRAPHICAL REFERENCES

TARI WEICHERDING, UNIVERSITY OF ILLINOIS

Management of geospatial data is an emerging challenge for academia, government agencies, and private companies. Several new geospatial data systems and related technologies are under development to address such changes.

One such program is the Alexandria Digital Library Project, which is one of six projects funded under the Digital Libraries Initiative (DLI), a joint program of the National Science Foundation, the Defense Advanced Research Projects Agency (DARPA), and the National Aeronautics and Space Administration. The Alexandria Project, which is based at the University of California, Santa Barbara (UCSB), brings together a unique blend of researchers, developers, and educators, spanning the academic, public, and private sectors. The goal of this project is to build a distributed digital library (DL) that allows users to access and manipulate information in a variety of classes of collection items in terms of geospatial reference.

The Alexandria Project's research and development accomplishments achieved as of July 1998 include

- The development and implementation of a new, three-tier architecture for ADL, including a Java client, HTTP middleware involving five standard interfaces to allow the easy construction of new clients, and two heterogeneous databases (catalog and gazetteer)
- Redesign of the user interface based on input from a series of evaluation studies that includes context-sensitive help
- Incorporation of instrumentation technology into the testbed to support and facilitate evaluation
- Incorporation of significant collections into the testbed
- Interfacing ADL with various external applications, such as a computational modeling system, image processing systems, and a collaborative desktop environment, using wrapper technology
- Agreements with the California Digital Library (CDL) to incorporate ADL in the first release of CDL before the end of 1998
- Numerous research contributions in the areas of access to information by georeference, distributed database support, image processing and access, high-performance and parallel computing support, and user evaluation (http://www.alexandria.ucsb.edu/).

A geospatial information research team working on the Alexandria Project, comprising Goodchild (leader), Carver, Geffner, Hill, Kemp, Kothuri, Larsgaard, and Smith, is responsible for investigating a variety of issues relating to the integration of spatially referenced information objects into ADL. Because of its focus on geographic data, ADL has encountered and addressed several issues that arise in this context. The transition to a digital world has significant impact on many of the conventions associated with geographic data, on arrangements for data production and dissemination, and on the ways in which data are used. During the past year, ADL has made significant progress on many issues such as the level of geographic detail, fuzzy footprints, the geolibrary, support for geocomputation, and libraries as central services.

First of all, many geographic phenomena are almost infinitely complex, so any attempt to capture them in geographic data must involve approximation. The level of detail of a data set is an important indicator of its volume, and thus of many practical issues of storage and dissemination. Within ADL, level of detail is a major determinant of the problems the user will encounter in attempting to decide whether a dataset meets the user's

needs, and of the time it will take to download. Browse images are provided in ADL to help overcome these problems, and ADL's work on wavelet decomposition and progressive transmission is also aimed in this direction.

Second, many user queries will be expressed in terms of geographic areas that do not have geographic referents, and are thus retrievable using geographic search mechanisms, but for which the corresponding footprints are fuzzy. Several issues must be dealt with if fuzzy footprints are to be incorporated into digital spatial data search and retrieval mechanisms. The Geospatial Information Research Team has experimented with three methods of formal representation of fuzzy footprints: a crisp polygon degraded by a simple mathematical function, a radially symmetric function, and a raster representation of a general surface. Appropriate methods were examined for eliciting such representations from users, both before and during the search process. A number of display methods have been implemented, in an effort to find methods that are as informative as possible to the user. Finally, the research team experimented with metrics of the goodness of fit between fuzzy or crisp representations of the user's area of interest, and fuzzy or crisp representations of an information object's footprint. It has been assumed throughout that fuzzy footprints can be described by one of a set of simple models, and there may be instances where none of these models is satisfactory. The approach assumes that a single model is appropriate, but there will certainly be instances where one group's concept of a fuzzy region differs from that of another group.

Third, it is physically impossible to build a geolibrary, although conventional map libraries come as close as it is possible to come. In a digital world, however, these problems disappear. The user of a digital geolibrary can be presented with a globe, can zoom to the appropriate level of detail, can access lists of placenames and see their footprints, and can move up or down the placename hierarchy using links between places. Moreover, a digital geolibrary solves the problem of physical access, if the services of the library are provided over a universal network like the internet. And finally, the collection of a geolibrary can be dispersed—a digital geolibrary can consist of a collection of servers, each specializing in materials about their local regions. The contents of the geolibrary would be very different from those of a conventional physical library. They would be dominated by multimedia information of local interest, in fact precisely the kinds of information needed by an informed citizenry, and one that is deeply involved in issues affecting its neighborhood, region, and planet. Because its contents would be different, a geolibrary might attract an entirely new type of library user.

Fourth, geocomputation has a large appetite for data. It focuses on modeling processes on geographic landscapes that can be sharply differentiated. In short, geocomputation, with its extensive data demands, is arriving as a novel paradigm at a time when many traditional arrangements for production and dissemination of geographic data are breaking down and are being replaced by a much more flexible, localized, autonomous, and chaotic system that is at the same time much richer, with far more to offer. While new technology has made far more data available, it has also created massive problems in making effective use of its potential. Paradoxically, only the technology itself can provide the basis of solutions.

Finally, libraries fall into the category of central facilities serving a dispersed population. The transition to digital information handling is in the process of engendering changes in many aspects of the central facilities model, including access (transition from physical access and delivery of media to access through electronic networks and delivery of bits), economies of scale (physical libraries replaced by digital servers), and consumer behavior (consumers have increasing numbers of choices). The future map of research libraries will look very different from today's. Instead of the classical pattern of central service provision, it will be sufficient for each information-bearing object (IBO) to be available from only a small number of servers, and under perfect connectivity, from only one. A research library will be able to focus on serving only those IBOs that are of particular relevance to its local role. Its responsibility to a geographically defined constituency

argues for it to serve those IGDIs (IBOs of geographically determined interest) whose footprints overlap its domain, or to provide indirect links to the respective custodians. The library's responsibility to its scholars argues for it to serve the results of their research and their contributions to the corpus of human knowledge, or to provide indirect access to IBOs on each scholar's personal server (though it will likely be argued that the institution is more persistent than the location of the scholar). The library may also serve IBOs that are of particular relevance to the interests of its scholars, or collections of archival IBOs that are analogous to today's special collections. In all other cases, however, the institution will rely not on its own library but on services provided collectively and paid for collectively. The research library of the digital world will be a much more specialized entity, reflecting the effectively infinite range and zero threshold of library service provision in the digital world (http://www.alexandria.ucsb.edu/).

The Alexandria Digital Library Project is utilizing several databases, including both Informix/Illustra and Oracle. Illustra DBMS (now owned by Informix Software, Inc.) is the primary catalog DBMS for the WP. Currently, Informix serves as the full catalog and gazetteer with Illustra version 3.2 on Solaris, and a smaller California-only catalog and gazetteer with Illustra version 2.4 on digital UNIX. The 6M-record gazetteer is one of the largest spatial databases supported by Illustra. Oracle Corporation is the world's leading supplier of software for information management, and the world's second-largest software company. Oracle, as part of the ADL, will leverage its expertise in Very Large Databases (VLDBs) as well as its ability to store and access large amounts of geographic information to assist in the creation of very large and complex databases. Oracle Universal Server with the Spatial Data Option will be used to build a distributed VLDB of geospatial information. The Spatial Data Option enables spatial information to be stored, accessed, and manipulated in the same manner as structured data and allows complete integration of complex dimensional data into the Oracle7 release 7.3 database. Spatial data can be geographic, scientific, demographic, physical, or time-referenced.

Some of the major plans for the Alexandria Project during the next six months include:

- Conversion of the ADL middleware to Java
- Redesign of the ADL catalog
- Develop an ingest system for the new gazetteer
- Evaluate the new Java-based client interface
- Integration of research results into testbed
- Making ADL an operational component of CDL
- Final user evaluation and testing.

Besides the Alexandria Project and the databases that the ADL uses, other companies and universities are dealing with geospatial data management. Core Software Technology is one company that develops and provides software and services to handle geospatial (image, cartographic, and demographic) data needs. The company's premier product, TerraSoar 3.0, is the first comprehensive solution for the distribution of geospatial data sets. Core's products take strategic advantage of technological progress and are based on open industry standards. Core also operates ImageNet, the leading commercial on-line geospatial visual indexing and distribution service. Through the company's ImageNet site on the World Wide Web, a user can simultaneously access numerous geospatial databases operated by Core and its affiliates across five continents.

TerraSoar allows an organization to integrate its geographic data into an enterprise-wide solution. TerraSoar utilizes the Digital Chart of the World to specify an area of interest and allows a user to zoom in to define his area of interest. TerraSoar is currently used by organizations to manage databases ranging in size from a few thousand records to over 50 million (http://www.coresw.com/CST/).

Using ImageNet, another company dealing with geospatial data management, a user

can view nearly all commercially available geospatial data associated with a geographic location. ImageNet responds by initiating a parallel search of networked databases worldwide and returns a collated list of databases with relevant information (http://www.coresw.com/CST/).

Core Software Technology (CST), the premier provider of geospatial data management solutions, announced its support for Oracle8i, the world's first database for Internet computing in November 1998 at Oracle OpenWorld. With the integration of CST's TerraSoar Web-based querying software and Oracle8iSpatial, the leading technology for spatial data management, users will be able to perform complex queries accessing distributed geographic data located over any region of the world. Users also can seamlessly integrate their spatial data into enterprise applications and fully leverage the scalability, reliability, and performance of the Oracle8i database. The combined solution will also allow users, over the Internet, to access any type of data sets ranging from text and sound files to video and satellite imagery with a single query (http://www.coresw.com/CST/).

In addition to the efforts of CST, Oracle Corp. announced in January 1998 that it was designating the University of Arkansas Center for Advanced Spatial Technologies (CAST) as its first Center of Excellence for Spatial Data Management. As a member of Oracle's Academic Alliance Program, CAST will integrate Oracle products into its existing academic curriculum and develop new curriculum featuring the Oracle8 Spatial Cartridge, Oracle's software technology for managing geospatial data.

The Center of Excellence for Spatial Data Management will help define the next generation of spatial applications and will provide students with the experience and knowledge to address the complex spatial data management issues government and private industry face today. In addition, Oracle's Spatial Cartridge product team plans to work with CAST to develop short courses, seminars, and professional development classes, an Oracle graduate assistantship tract, and a summer internship program on-site at Oracle.

According to Fred Limp, director of CAST, "We have decided to shift all our database applications and research from Informix to Oracle." The decision has been based primarily on Oracle's development of the Spatial Cartridge technology and the way in which they have integrated geospatial data into their object-relational database management system (RDBMS) and their support for the OpenGIS standards. It is clear that object-relational DBMS provides substantial advantages over existing spatial data storage systems, and Oracle has the best geospatial DBMS. Their Spatial Cartridge is a part of core Oracle technology and Oracle is strongly committed to geospatial data—it is not a third-party add-on to their system (http://techmall.com/).

APPENDIX G: USACE SAMPLE NETWORK ARCHITECTURE AND SAMPLE ORGANIZATION NETWORK ARCHITECTURE

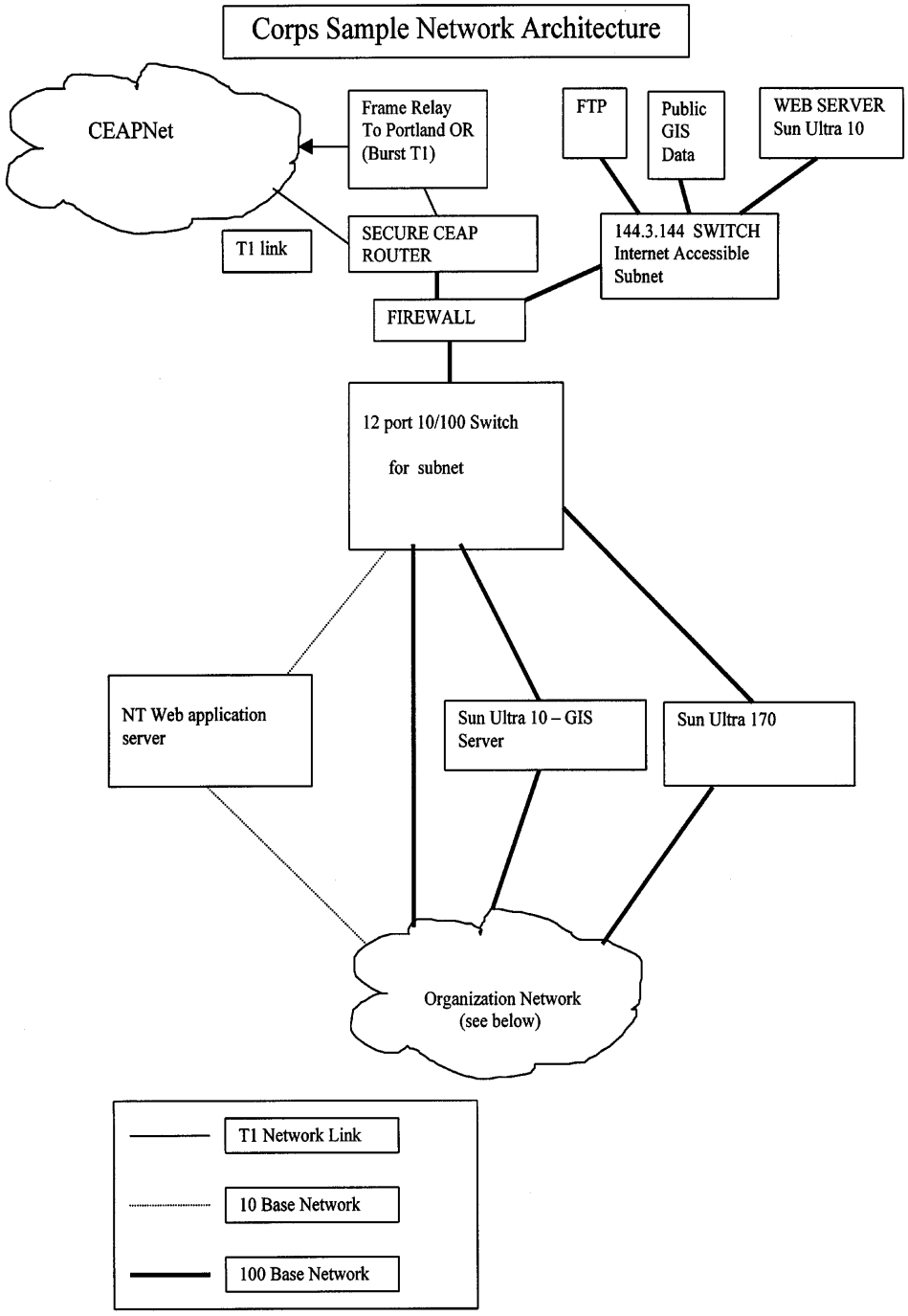


Figure G1. An example of current USACE network architecture.

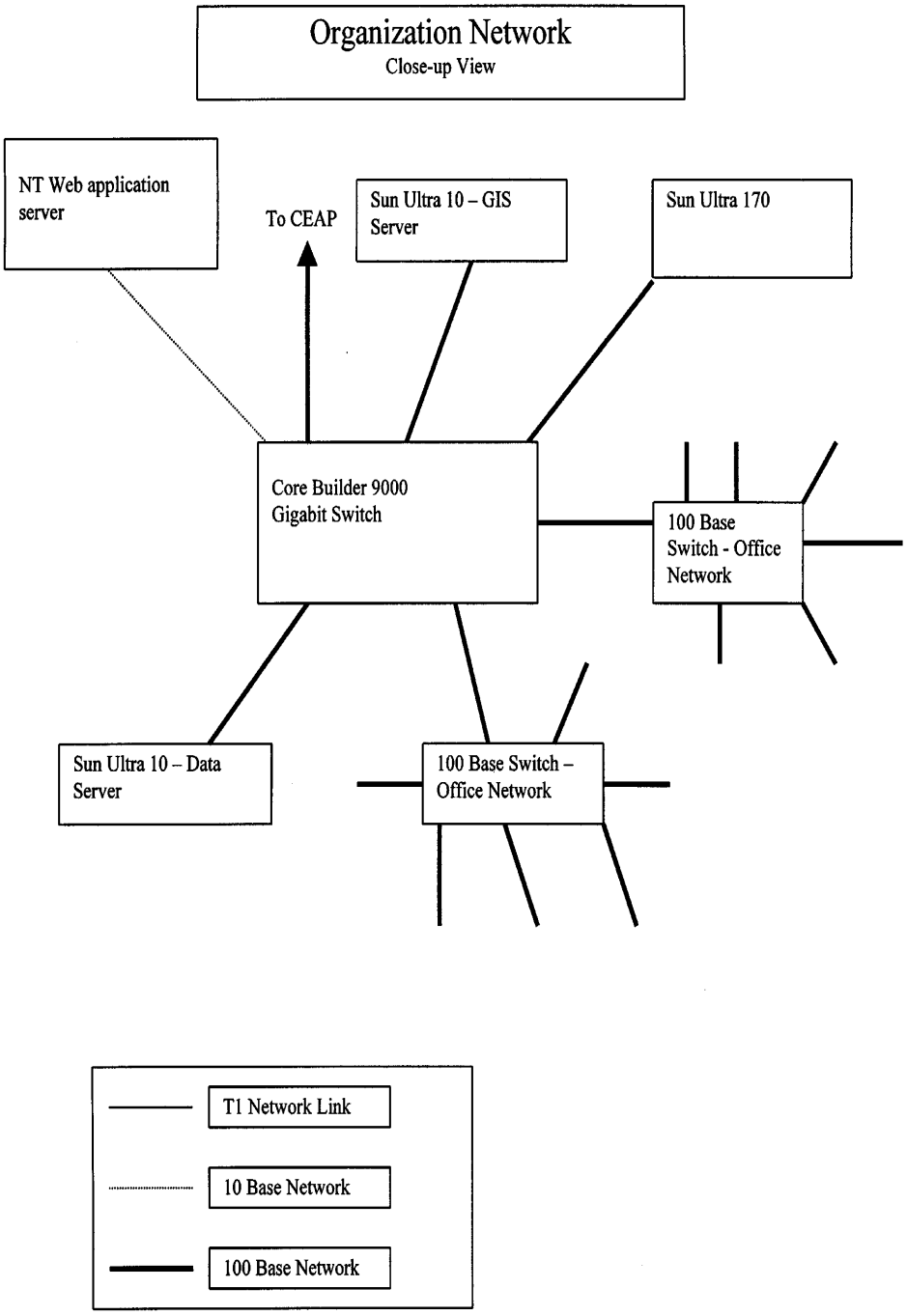


Figure G1 (cont'd). An example of current USACE organization network architecture.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
November 2000	Technical Report	

4. TITLE AND SUBTITLE

U.S. Army Corps of Engineers Geospatial Data and Systems Management

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Nancy H. Greeley, Kelly M. Dilks, Chad M. Adams, and Gregg C. Hoge

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
33172

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Engineer Research and Development Center
Information Technology Laboratory—Hanover
72 Lyme Road
Hanover, New Hampshire 03755-1290

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDC TR-00-9

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR / MONITOR'S ACRONYM(S)

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

Available from NTIS, Springfield, Virginia 22161.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

To help the U.S. Army Corps of Engineers manage its substantial investment in geospatial data and systems (GD&S), a review of the current state of GD&S in the Districts and Divisions was needed. The authors surveyed employees involved in GD&S work in multiple functional areas at multiple sites. The survey contained questions on GD&S administration and personnel issues including data management, GD&S maintenance, data libraries, data sharing, and Information Management involvement with GD&S. Other topics areas covered are GD&S workflow and technical issues, including hardware, software, metadata creation and publishing, data storage and distribution, networks, and security. The results of the survey show great differences in the approaches and challenges within the various Districts and Divisions. Many are finding solutions that are shared in this report. The report is written as a manual for managers. Each existing problem is described followed by possible solutions. Seven conclusions and a summary "Manager's Checklist" of GD&S issues are included.

15. SUBJECT TERMS			
	GD&S	Geospatial data and systems	Geospatial data sharing
	Geographic information systems	Geospatial data libraries	Geospatial metadata
	Geospatial data	Geospatial data management	Geospatial systems

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	U	50	