AIIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910

301/587-8202

Centimeter

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 mm

1 2 3 4 5

Inches

MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

1 of 1

Conf-9409190--1

WSRC-MS-94-0394

Using Geographic Data Browsers in a Networked Environment (U)

by

H. E. Mackey

Westinghouse Savannah River Company
Savannah River Site
Aiken, South Carolina 29808

P. T. Bresnahan
University of South Carolina
SC USA

D. J. Cowen
University of South Carolina
SC USA

G. B. Ehler
University of South Carolina
SC USA

E. King
University of South Carolina
SC USA

A document prepared for PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON SPATIAL DATA HANDLING at Edinburgh from 09/08/94 - 09/08/94.

DOE Contract No. DE-AC09-89SR18035

DISCLAIMER

WSRC-MS-94-0394

USING GEOGRAPHIC DATA BROWSERS IN A NETWORKED ENVIRONMENT

by

Partick J. Bresnaham, David J. Cowen, Geoffrey D. Ehler, Eddie King, W. Lynn Shirley, and Timothy White
Humanities and Social Sciences Computing Laboratory
University of South Carolina 29208

This information contained in this article was developed during the course of work under Contract No. DE-AC09-89SR18035 with the US Department of Energy.

Using Geographical Data Browsers in a Networked Environment

Patrick J. Bresnahan*, David J. Cowen, Geoffrey B. Ehler, Eddie King,
W. Lynn Shirley, and Timothy White

Humanities and Social Sciences Computing Lab
University of South Carolina
Columbia, South Carolina 29208
pat, cowen, geoff, eddie, lynn, tim@otis.hssc.scarolina.edu

Abstract. The rapid development of spatial information technology has promoted a large variety of applications. The use of diverse datasets in these applications has often been limited to single-platform analysis and distribution. Recent improvements in networking capabilities have led to the possibility of cross-platform access to geographical information and improved interoperability. To promote access to various databases and limit unsecured use of datasets across networks, spatial data browsers have been developed to allow simple analysis of data void of permanent update. Furthermore, applicable network configurations have greatly improved dissemination of views of existing data and individual use of distributed heterogeneous databases. Various user interfaces and networking applications can be combined to maximize data flow between storage media and data browsing software. This paper reviews the use of such distributed data browsing systems and the characteristics of platform-specific access and performance. Distinct discussions will review both geographical browser interface and networking services components. The selection of networking and software solutions was based on the development of a data browsing system for a twenty-user environment of mixed platforms (Unix, DOS, and Macintosh). Seventy percent of the user environment, however, was Macintosh-based.

Introduction

Development of computer network technology has encouraged increased utilization of spatial information. This increase is the result of expanding abilities to 'mix and match', merge, and access diverse databases with little effort, via networks (Camarata 1992, Handley *et al.* 1992). Additionally, interoperability has emerged as an important facet in today's computer industry. Networked multi-platform solutions have greatly advanced distributed computing to promote interoperability. The proliferation of microcomputers in Unix-dominated environments demands uniformity in access methods. With this in mind, operating system developers are migrating Unix-like network tools to microcomputer platforms. Microsoft's Windows NT and IBM's OS/2 operating systems (OS) contain examples of this migration. Simultaneously, hardware vendors are experiencing a similar transformation as more processors are designed with increased flexibility. The cooperative design of the Power PC between Apple and IBM is the most noteworthy example of this trend.

Unfortunately, increased demands for data access complicates interoperability parameters and leads to an escalation of data security problems. To this end, data browsing interfaces have been developed on top of networking services to access 'views' of existing data over networks while offering the ability to perform simple analysis and output functions. As discussed by Abel *et al.* (1992), views are representations of existing data that are displayed by a user interface which allows spatial query of specific features. These customized displays of graphic elements may be output in various file formats. The view may also be used to access attribute tables which, in turn, may be output in text or ASCII formats to be included in further analysis by specific users.

How do geographic information systems (GIS) fit into these industry trends? GIS practitioners are primarily concerned with the relationships between large quantities of spatial and tabular data. At the same time, computer performance is critical in many GIS operations. These two issues can be balanced in a number of ways when a proper computing environment is established for any organization. In a truly heterogeneous spatial data handling system, users should be able to select desktop computing platforms on a task-specific basis. Furthermore, centralized or distributed databases should be maintained in a manner such that custodial duties and user access are not in conflict. Oftentimes, the ramifications of these decisions go far beyond organizational boundaries. The failure to provide a proper foundation for data access will lead to unnecessary duplication of data.

Data browsing applications that are included in networked systems increase the availability of existing data from various sources. Layers of information from GIS, remotely sensed images, and a variety of multi-media data may be accessed from numerous servers and used on a single platform. However, contemporary systems and networking standards do not allow simplistic seamless flow of data between platforms. As a result, networking solutions must be developed for each operating system until future standards eliminate such barriers. A variety of solutions exists for current implementation. Custom X Windows-based interfaces have been developed to utilize networking applications for both browsing and analysis of data located on dedicated central servers (Matteson *et al.* 1993, McBride *et al.* 1993). Such systems are designed to serve project-specific needs and may be part of database development initiatives which would be difficult to duplicate in other locations. Similar systems also utilize X solutions as limited data browsers, based on spatial metadata, that may be duplicated in other locations but are limited by native transfer protocols for use with X11 Window Systems on Unix platforms requiring similar clients (Evans *et al.* 1992).

Multiple or distributed servers require more flexible interface applications and network configurations to transfer views of data from autonomous sources with increased heterogeneity. The increase in system flexibility also promotes platform interoperability. To this end, networking and user interface technology must allow integration of databases without physically centralizing servers (Newton *et al.* 1992). Network connection to file servers must promote a level of interoperability where data, not the user, are passed through the network to individual platforms in their native format (Calvo and McDonald 1993). Development of interface and networking solutions is critical to the advancement of spatial data agendas concerning information access and dissemination as described in plans for the U.S. National Spatial Data Infrastructure (NSDI). Improved spatial data browsing and transfer would allow access to all levels of research via centralized indexes of distributed files within the NSDI (Tosta, 1992). The significance of such developments may be measured by the recovery and use of "silos filled with rotting data," or exformation, via interface and network applications using the NSDI (Gore, 1991).

Geographical Browser Interfaces

Important elements of browsing geographic data over a network include locating, viewing, and accessing and/or retrieving files. Additionally, user interfaces must portray accurate views of both complex graphic and large amounts of attribute data due to the uniqueness of spatial data. Unlike other forms of remote database access, geographical browsers in a networked environment provide simple GIS analysis of remotely stored data. Many commercial data browsers handle these tasks efficiently with additional analysis and output features.

The most sophisticated browsing applications allow both spatial and logical expression queries to locate files of interest. Hyper-text technologies are being added to locational functions to provide guided movement through file directories, thus simplifying searches by providing options. An example of this type of access may be seen in hyper-text macro language (HTML) interfaces, such as Mosaic and WAIS/WWW clients, currently being developed for multiple operating systems (Armenakis 1993). However, locating data is usually inhibited by the examination of hierarchical directories that are common on all platforms. With this type of standard access, users depend on descriptive directory and file naming conventions for guidance.

Once data is located, browsing software may display the complete file or may present a simplified view via generalization routines or metadata (Evans *et al.* 1992). Inherently, those applications which utilize generalized views have limited analysis functionality. These applications are often designed to be data locating tools for subsequent downloading of selected files (Schramm 1993). More robust analysis may be performed on actual data residing on servers which is addressed by the browser through directory path assignments. This, of course, relies heavily on fluid movement of data over the network but limits the amount of data duplication at separate sites. Often, the network traffic obstacle may be diminished by saving data in local random access memory (RAM) to speed successive display and analysis. Datasets that are accessed frequently may also be copied to local disk to reduce network interaction as permissions dictate.

The significance of geographical data browsing, as opposed to more static applications, is the dynamic ability to handle complex graphic elements while accessing large attribute tables. The analytical functions of these browsing applications are usually limited to spatial searches and structured tabular queries. Resulting cartographic and tabular output may be incorporated into various software for use in reports, maps, and visual presentations. The utility of geographical browsers that access networked databases may be seen in the increased interest in heterogeneous spatial data supporting planning, design, and decision-making processes (Newton *et al.* 1992).

Network Services

The functionality of data browsers is dramatically increased when networked file servers are accessible. Database server features of networked systems limit data redundancy and maintenance while making efficient use of disk space. Aside from the benefits gained by the deposition of data in centralized locations, it is also advantageous to allocate the geographical application software from these servers. This facilitates upgrades to the applications by providing limited points of maintenance. As discussed by Coleman and Zwart (1992), the advantages of high-speed spatial information networks include: 1) improved browsing and downloading of image and graphics files at remote sites; 2) routine update from branch databases by a centralized management staff; 3) and the ability to run complex modelling functions on more powerful workstations using data from remote sites.

There are a number of trade-offs involved in configuring any particular system. Of prime importance is the processing power of the "local" machine versus the server, the number of users on the network, and the configuration of the network. One common scenario would consist of a handful of users on a network in which processing is executed on a powerful server (Figure 1). In this situation, an X-windows application would provide a productive work environment for multiple users with minimal desktop computing power. Another common scenario would require processing to be carried out at individual workstations with data storage on a central disk (Figure 2). Any alteration of these factors would tempt users to gain control of essential resources. This may evolve to the point where user network access is reduced to printing and plotting functions. The result is usually that of wasted resources. Unfortunately, this has become more common as disk prices have declined, more useful and stable data are distributed on CD-ROM, and network traffic has dramatically increased.

A critical link to balancing these components is the underlying networking hardware and protocols that provide the path for traffic flow in such heterogeneous environments. Each user environment has its own native protocol, but the ability to mix these protocols is crucial. Transmission Control Protocol/Internet Protocol (TCP/IP) provides network portage for Unix platforms while AppleTalk is utilized by Macintoshes and Intel-based machines employ Internetwork Packet Exchange (IPX) or Network Basic Input Output System (NetBIOS) (Derfler and Freed 1993). Ideally, each workstation should operate from its native protocol. The technology is available, however, to allow most platforms to function using the native protocols of other platforms.

	TCP/IP	**AppleTalk**	**IPX or NetBIOS**
UNIX	*native*	available	available
Macintosh	available	*native*	available
Intel-based PC	available	available	*native*

Some of these options would be more desirable than others depending on the specific site implementation. Also, the availability of some combinations may be limited for specific hardware/operating systems. The most logical way to implement such a system would allow each workstation to operate in its native form with protocol support provided at some centrally-administered machine (Figure 3). The simultaneous need to centralize certain data makes this more convenient with respect to hardware, software, and data maintenance.

To demonstrate the effectiveness of browsing geographical data using platform specific-software, a limited test was performed using an existing production environment. In this environment, each platform used its native networking protocol as described above. The most obvious solution in this environment was to centralize data storage on an existing Novell Netware server. From the server, IPX, AppleTalk, and TCP/IP were accessible. By requiring the server to support all three protocols, client workstations could be used as configured for their normal daily tasks (Figure 4). For comparison, the same functions were attempted with the data stored on local disks. The characteristics of each machine are outlined in Table 1.

Although the geographical browsing software used in testing does not currently support transmission rate benchmarking, a relative test was performed among platforms using dual hand-held stopwatches. The objective was to identify elapsed time between keyboard selection of items for display and completed display, thus, in effect, measuring data access time and graphic display rates. Sample selections included a multispectral image, a vector-based map layer, and a dual selection query (SQL). The 17.4 megabyte (mb) multispectral image was acquired in ERDAS format (.lan) while the 3.3 mb GIS data layer (containing 378,375 arcs) was accessed in Arc/Info format. The Boolean query contained 3940 records. Although the method might seem crude, the user-perceived rates of data acquisition and display were relatively useful in identifying basic characteristics of network speed.

All three data types were queried (or drawn) five successive times using both the Netware server and the local disk. Those results were averaged for each platform-data access combination. It was apparent that local RAM accounted for slight decreases in successive display times. The assumption, that local disk access speeds would be faster, was also substantiated. However, to efficiently utilize network components, local disk storage should be kept to a minimum.

To summarize the results, all three platforms demonstrated similar results retrieving each data type from the server (Table 1). Disk access rates from the server varied from 1.6% to 49.5% slower than local disk access for Unix and Macintosh platforms. The Intel-based machines, however, demonstrated retrieval times approximately 2-3 times slower using the server as opposed to local disk access. This response is suspected to be the result of an extremely fast local disk combined with the OS/2 operating system. Overall, both Macintosh and Intel machines provided comparable performance with respect to the Unix platform with two exceptions. The Macintosh tested significantly slower when retrieving the vector file from the server while raster data slowed the Intel-based machine.

The level of performance for each platform is quite promising for native-based network protocols. The use of different operating systems by clients is well supported by favorable data transfer and display rates. The slower response of Intel-based machines does, however, cause concern in this particular browser/network protocol environment. As discussed earlier, this is just a single solution in the array of configurations which must be designed to meet user-specific goals. These goals are often defined by an organization's information management strategy and the activities that are conducted using available resources (Coleman and Zwart 1992).

Summary

Developing networking technology has, indeed, promoted increased access to diverse geographic databases. However, autonomous interfaces and networking protocols have hindered the seamless flow of data between archives and users. This discussion was intended to review several solutions to current roadblocks while emphasizing the complexity of system-dependent possibilities. The 'proof-of-concept' example of using native protocols for data transfer was discussed to portray a single configuration that promotes browser interface-independent integration of data resources. The goal of future hardware and software developers should be to design systems that implement this type of user-defined environment void of protocol limitations.

Contemporary browsing interface and networking technology promotes numerous nodes of access. However, more complex solutions have lead to expanding control problems. Until each of the connected nodes of the system is totally controlled, issues of compatibility and performance will continue. To this end, standardization of interface and networking protocols should be emphasized to promote uniformity in application and network implementation. As systems become less protocol dependent, use of autonomous distributed and centralized databases should increase. This evolution should put geographic resources in the hands of the user.

* This research was supported, in part, by an appointment to the U.S. Department of Energy Laboratory Cooperative Research Training Program at the Savannah River Technology Center administered by the Oak Ridge Institute for Science and Education.

Support was received from the South Carolina Universities Research and Education Foundation under Task Order #111 which was administered by Halkard E. Mackey of the Westinghouse Savannah River Company.

References

Abel, D. J., Yap, S. K., Ackland, R., Cameron, M. A., Smith, D. F., and Walker, G., 1992, Environmental decision support system project: an exploration of alternative architectures for geographical information systems. *International Journal of Geographical Information Systems*, **3**, pp. 193-204.

Armenakis, C., 1993, Hypermedia: an information management approach for geographic data. In *Proceedings of GIS/LIS '93, held in Minneapolis, Minnesota in November, 1993* (Bethesda, Maryland: American Congress on Surveying and Mapping and the American Society for Photogrammetry and Remote Sensing), **1**, pp. 19-28.

Calvo, S. C., and McDonald, K. R., 1993, Accessing Distributed Heterogeneous Earth Science Inventories via the EOSDIS Version 0 Information Management System. In *Technical Papers of the ACSM/ASPRS Annual Convention & Exposition, held in New Orleans, Louisiana in February, 1993* (Bethesda, Maryland: American Congress on Surveying and Mapping and the American Society for Photogrammetry and Remote Sensing), **3**, pp. 48-55.

Camarata, S. J., 1992, The Integration of New Technologies and Distributed Architecture: a new revolution for geographic information systems. In *Networking Spatial Information Systems*, edited by P. W. Newton, P. R. Zwart, and M. E. Cavill (London: Belhaven Press), pp. 49-62.

Coleman, D. J., and Zwart, P. R., 1992, Modelling Usage and Telecommunications Performance in Real-time Spatial Information Networks. In *Proceedings of the 5th International Symposium on Spatial Data Handling, held in Charleston, South Carolina in August, 1992* (Columbia, South Carolina: International Geographical Union, Commission on GIS), **1**, pp. 144-153.

Derfler, F. J., and Freed, L., 1993, *How Networks Work* (Emeryville, California: Ziff-Davis).

Evans, J. D., Ferreira, J. Jr., and Thompson, P. R., 1992, A Visual Interface To Heterogeneous Spatial Databases Based On Spatial Metadata. In *Proceedings of the 5th International Symposium on Spatial Data Handling, held in Charleston, South Carolina in August, 1992* (Columbia, South Carolina: International Geographical Union, Commission on GIS), **1**, pp. 282-293.

Gore, A., 1991, Infrastructure for the Global Village: a high capacity network will not be built without government investment. *Scientific American*, September, pp. 150-153.

Handley, T. H. Jr., Li, Y. P., Jacobson, A. S., and Tran, A. V., 1992, DataHub - Knowledge-based Science Data Management. In *Technical Papers of ASPRS/ACSM/RT 92, held in Washington, D. C. in August, 1992* (Bethesda, Maryland: American Congress on Surveying and Mapping and the American Society for Photogrammetry and Remote Sensing), **3**, pp. 122-134.

Matteson, W. H., Wayne, L. D., Carpenter, M., McBride, R. A., and Williams, S. J., 1993, Louisiana Coastal GIS Network: A Cataloging Framework For Spatial Data. In *Proceedings of GIS/LIS '93, held in Minneapolis, Minnesota in November, 1993* (Bethesda, Maryland: American Congress on Surveying and Mapping and the American Society for Photogrammetry and Remote Sensing), **1**, pp. 312-322.

McBride, R. A., Matteson, W. H., Jones, F. W., Braud, D. Jr., Wayne, L. D., Streiffer, H. R., Carpenter, M., Lewis, A. J., Mogli, S., Arnold, J., Lingineni, S., and Williams, S. J., 1993, Louisiana Coastal Geographic Information System Network: More Than a Spatially Indexed Cataloging System. In *Technical Papers of the ACSM/ASPRS Annual Convention & Exposition, held in New Orleans, Louisiana in February, 1993* (Bethesda, Maryland: American Congress on Surveying and Mapping and the American Society for Photogrammetry and Remote Sensing), **3**, pp. 238-255.

Newton, P. W., Zwart, P. R., and Cavill, M. E., 1992, Introducing the Next Wave: networking spatial information systems. In *Networking Spatial Information Systems*, edited by P. W. Newton, P. R. Zwart, and M. E. Cavill (London: Belhaven Press), pp. 1-11.

Schramm, W. G., 1993, NEONS: A database management system for environmental data. *Earth Systems Monitor*, Vol. 3, **4**, pp. 7-8.

Tosta, N., 1992, The National Spatial Data Infrastructure. *Geo Info Systems*, November/December, pp. 30-35.

Table 1

Test 1. Retrieval and Display of a 3.3 mb File of 3,940 Polygon Boundaries. Average time of five trials in seconds.

	Local Disk	Novell Server	Difference Time	Difference Percent
SUN Sparc 10	31.42	31.18	- 0.24	- .76
Macintosh, Quadra 800	37.59	49.55	+12.06	+ 32.00
IBM 80486,DX2/66	15.05	33.69	+18.64	+123.80

Test 2. Retrieval and Display of a 17.4 mb 24 bit image. Average time of five trials in seconds.

	Local Disk	Novell Server	Difference Time	Difference Percent
SUN Sparc 10	8.13	9.58*	+1.46	+17.80
Macintosh, Quadra 800	25.14	26.44	+1.30	+ 5.17
IBM 80486,DX2/66	10.43	50.98	+40.55	+388.78

* SUN workstation times were 21.57, 5.32, 5.20, 5.37, and 5.22 seconds. The significant reduction in time is related to use of cache memory. The initial retrieval time of 21.57 seconds is more likely during non-successive operations.

Test 3. Dual expression SQL Query and Retrieval of 3,940 records. Average time of five trials in seconds.

	Local Disk	Novell Server	Difference Time	Difference Percent
SUN Sparc 10	1.33	1.58	+ .25	+18.80
Macintosh, Quadra 800	4.86	6.71	+1.85	+38.07
IBM 80486,DX2/66	2.18	4.61	+2.43	+111.47

The specific machines used for the test were:

Novell server: IBM Model 95 with Intel 50 mhz 80486 DX processor and 64mb RAM;
SUN Microsystems Sparc 10 workstation with 32 mb RAM;
Apple Quadra 800 with 24mb RAM;
IBM Value Point with Intel 80486 66 mhz DX2 processor and 20 mb RAM.

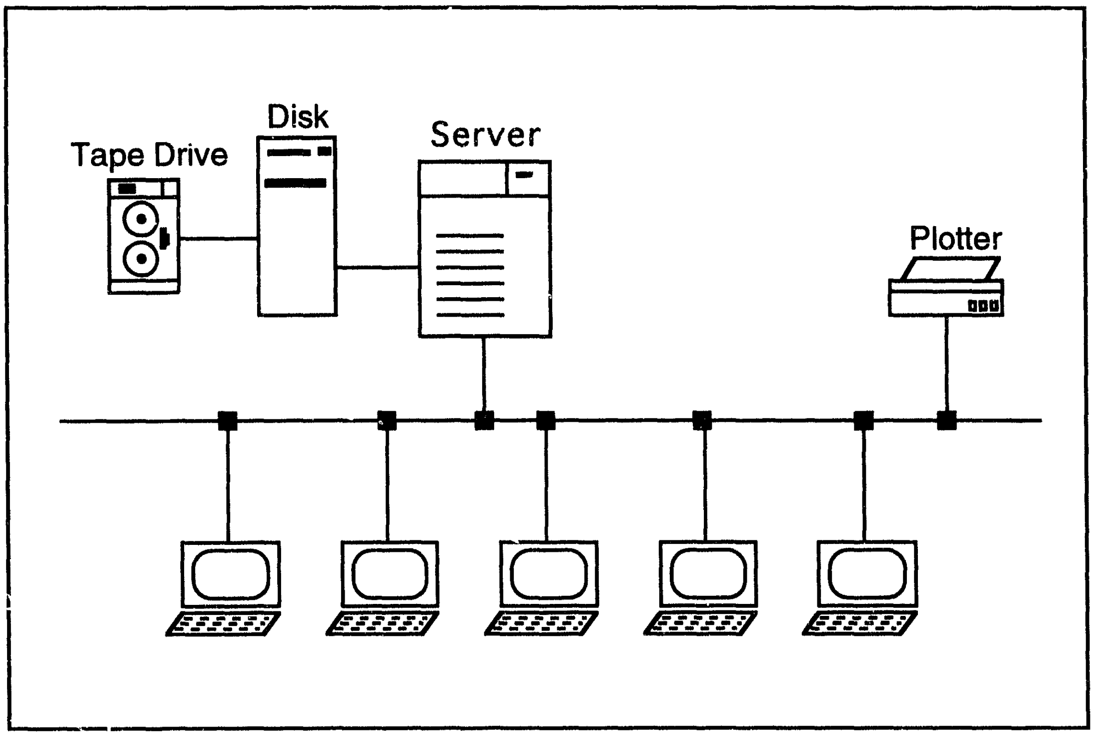

Figure 1 - Centralized processing in a network.

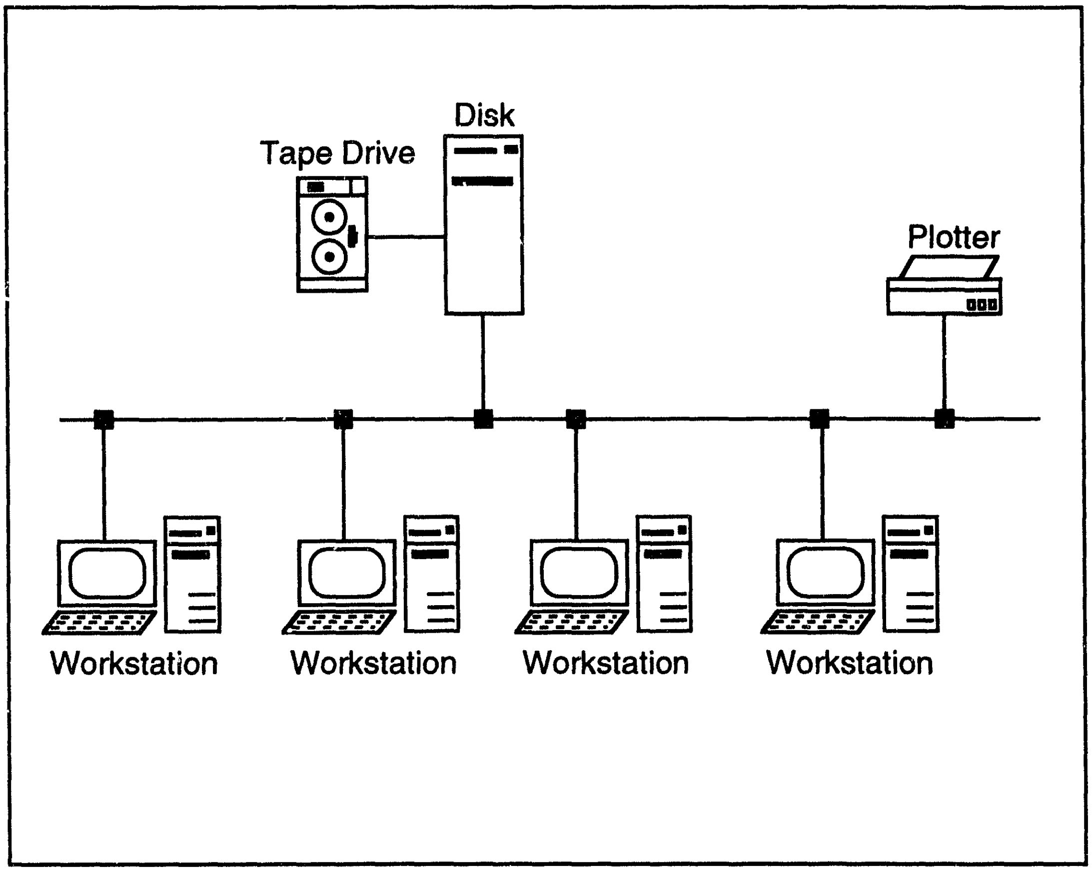

Figure 2 - Networked workstations utilize desktop power while accessing centralized databases.

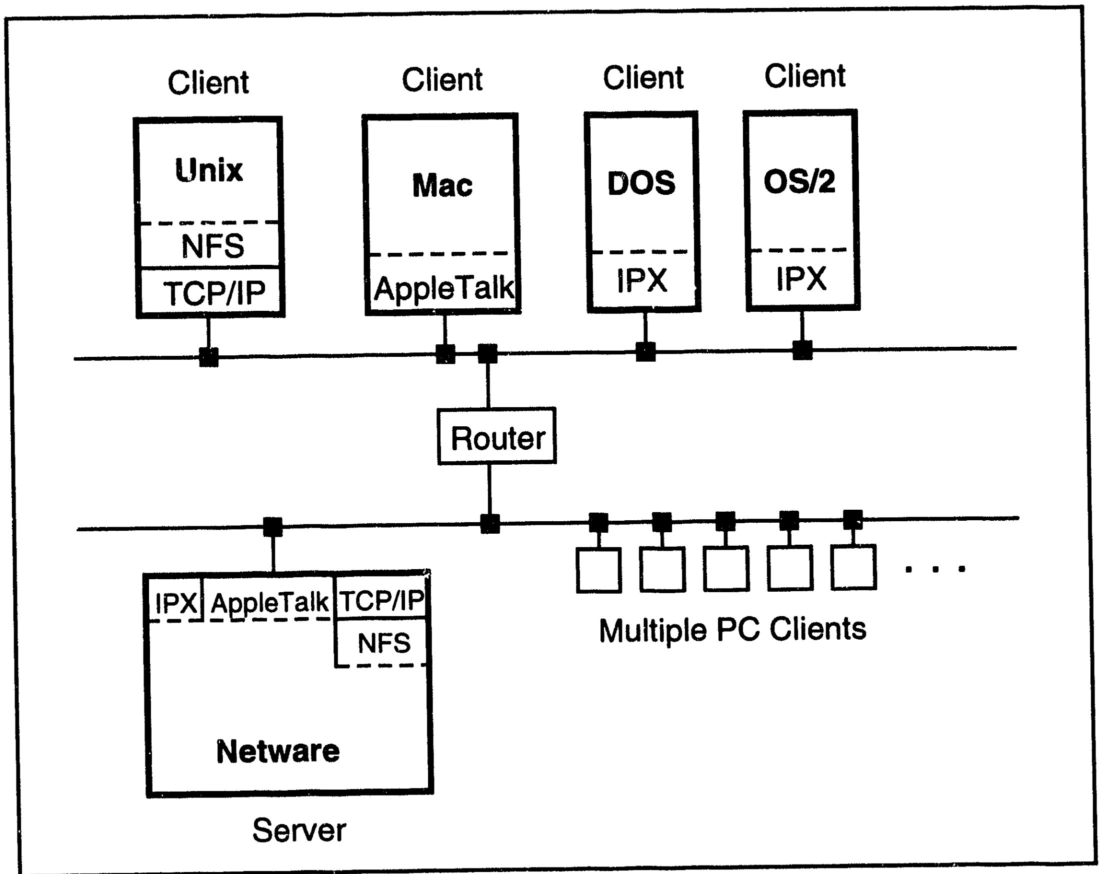

Figure 3 - Centralized protocol support allows each workstation to access data in native format.

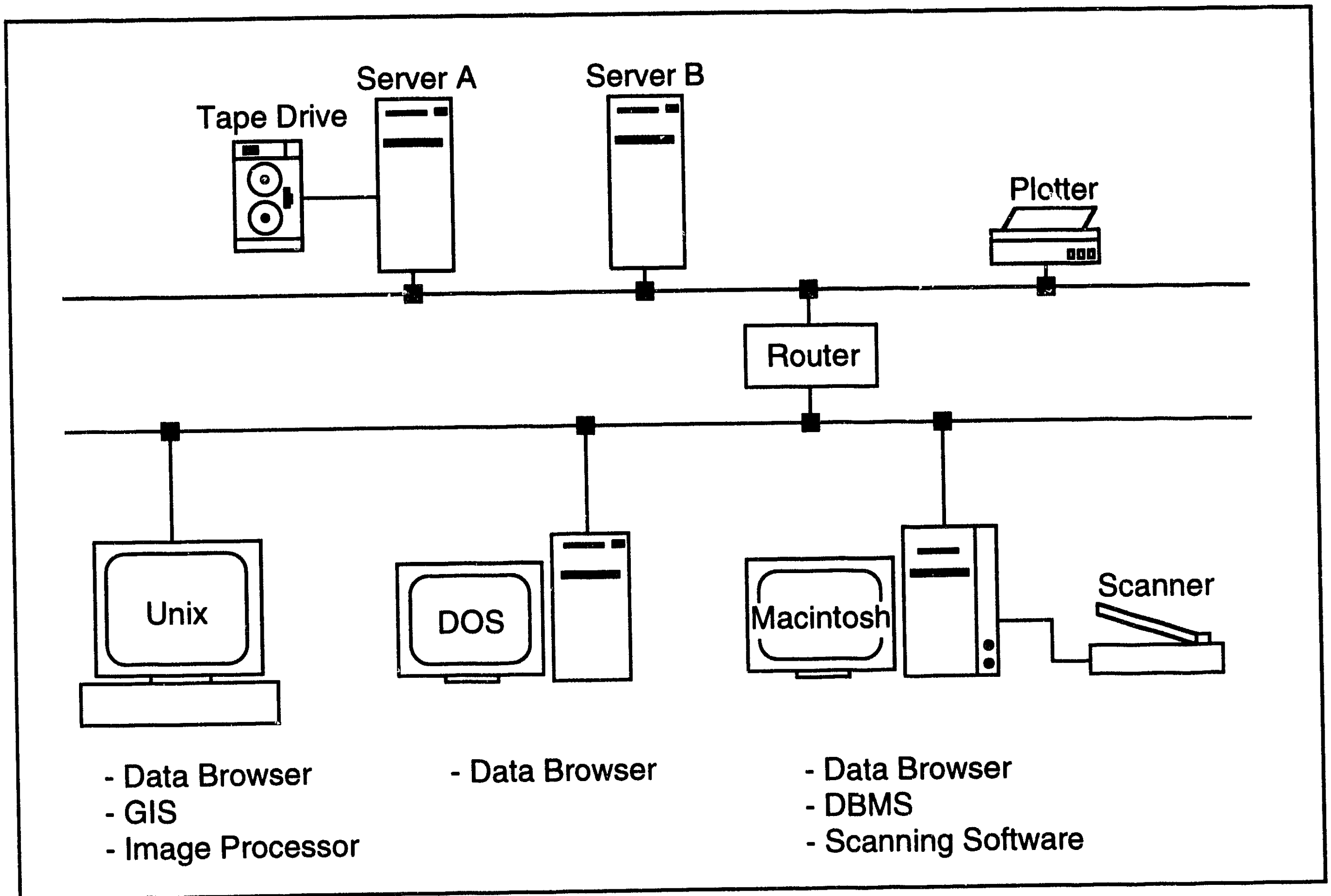

Figure 4 - Multi-platform access to geographical databases.

END

DATE FILMED

10/18/94