N89-26598

DEVELOPMENT OF AN INTELLIGENT INTERFACE FOR ADDING SPATIAL OBJECTS TO A KNOWLEDGE-BASED GEOGRAPHIC INFORMATION SYSTEM

William J. Campbell
National Space Science Data Center
Goddard Space Flight Center

Craig Goettsche
Science Applications Research Inc.
National Space Science Data Center
Goddard Space Flight Center

ABSTRACT

Earth Scientists lack adequate tools for quantifying complex relationships between existing data layers and studying and modeling the dynamic interactions of these data layers. There is a need for an earth systems tool to manipulate multi-layered, heterogeneous data sets that are spatially indexed, such as sensor imagery and maps, easily and intelligently in a single system. The system can access and manipulate data from multiple sensor sources, maps, and from a learned object hierarchy using an advanced knowledge-based geographical information system.

Such a prototype system, called KBGIS (Knowledge-Based Geographic Information System) was recently constructed at the University of California, Santa Barbara campus and funded by the United States Geological Survey. Although much of the system internals are well developed, the system lacks an adequate user interface, and would benefit from being able to input and output imagery data in NASA formats. The application of KBGIS would be of great benefit to current NASA earth science research as well as providing a more sophisticated understanding of the issues and required technologies in the upcoming EOS era.

This paper describes a methodology for development of an intelligent user interface and extending KBGIS to interconnect with existing NASA systems, such as imagery from the Land Analysis System (LAS), atmospheric data in Common Data Format (CDF), and visualization of complex data with the NSSDC Graphics System (NGS). This would allow NASA to quickly explore the utility of such a system, given the ability to transfer data in and out of KBGIS easily. The use and maintenance of the object hierarchies as polymorphic data types brings, to data management, a whole new set of problems and issues, few of which have been explored above the prototype level.

INTRODUCTION

It is a well known fact that the sheer volumes and complexity of scientific data being generated today require sophisticated technologies to locate, access, manipulate and display these data if they are to be of any significant scientific value. One technology that has come into its own is Geographic Information Systems (GIS) which has matured over the last twenty years such that it is now a multimillion dollar industry. GIS's allow almost any kind of data to be organized into a digital format such that any file, image or picture is transformed into a two-dimensional grid of points. For example, if a variable is an intensity function, say from a weather satellite, then this radiance value to percentage reflectance can be assigned a number expressing that intensity. Once this is accomplished all data now possess spatial singularities in that they may be referenced to a common geographical base. Usually what is done with geobased data is to reduce or enlarge and rectify the spatial attributes to one scale and projection. Now that the database is established in a commonly organized format, decision rules permit assignment of the relevant information from one or more classes that happen to fall within a given spatial sector or cell. This allows for a straightforward process for the user to compare data sets, produce new overlay combinations, assess the influence or interaction of different variables or map separates and provide input into models the user is trying to construct. This in essence is the "heart" of GIS data management. [CAM81]

Distinction between GIS and a Database Management System (DBMS)

"A Data Base Management System (DBMS) is a set of programs used for the manipulation and retrieval of logically related files containing data and structural information; it allows analysis to be used in some decision-making process. A GIS is a geo-referenced system for the specification, acquisition, storage, retrieval, and manipulation of data. These data may be related to a place; that is, the data elements link the data to a location identifier (e.g., Goddard Space Flight Center), whereas a DBMS does not require the location distinction". [CAM81] Figure1 is a graphical representation of a typical GIS data handling approach. As is apparent from Figure 1, a GIS should have at a minimum, the following capabilities:

- Data Input
- Encoding Procedures
- Data Management
- Manipulative Operations
- Output Products

Although GIS technology has come a long way since the mid 1960's, it still suffers from a variety of shortcomings such as a standardized, uniform interface, an accurate data capture and encoding mechanism, full integration with other related technologies, (e.g., image processing relational data models, graphics etc.). However, several universities and private vendors are taking a variety of approaches to address these shortcomings, one of which is described in the next section.

Artificial Intelligence and KBGIS

The logical fusion of Artificial Intelligence (AI) technologies to the complex, multifaceted capabilities of GIS's is an obvious one. Natural language query processors, expert systems, knowledge acquisition tools and artificial neural nets are but a few techniques that hold great promise for enhancing existing GIS shortcomings. One approach taken by the University of California Santa Barbara (UCSB) has been the establishment of a project to develop a Knowledge-Based Geographic Information

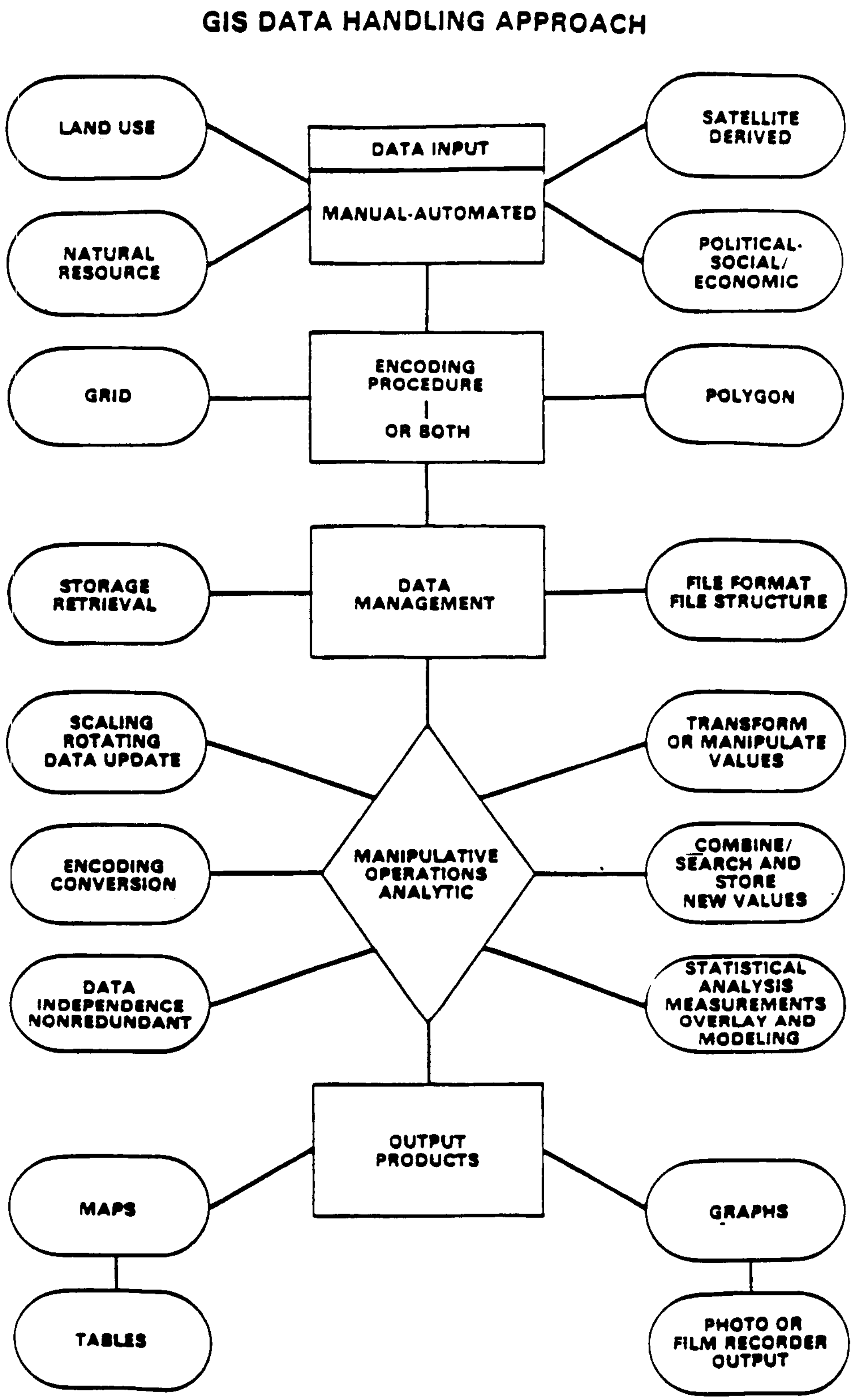
GIS DATA HANDLING APPROACH
LAND USE
NATURAL RESOURCE
DATA INPUT
MANUAL-AUTOMATED
SATELLITE DERIVED
POLITICAL-SOCIAL/ECONOMIC
GRID
ENCODING PROCEDURE | OR BOTH
POLYGON
STORAGE RETRIEVAL
DATA MANAGEMENT
FILE FORMAT FILE STRUCTURE
SCALING ROTATING DATA UPDATE
TRANSFORM OR MANIPULATE VALUES
ENCODING CONVERSION
MANIPULATIVE OPERATIONS ANALYTIC
COMBINE/SEARCH AND STORE NEW VALUES
DATA INDEPENDENCE NONREDUNDANT
STATISTICAL ANALYSIS MEASUREMENTS OVERLAY AND MODELING
OUTPUT PRODUCTS
MAPS
GRAPHS
TABLES
PHOTO OR FILM RECORDER OUTPUT

Figure 1.

System (KBGIS). Funded and supported by the United States Geological Survey and NASA, KBGIS is being developed to accommodate high-level expert system rules and heuristics, efficient spatial data search techniques through the use of new data structures and develop a learning capability by interaction from a user. The basic difference between a knowledge-based GIS from traditional GIS's or for that matter DBMS's is that a Knowledge-based GIS embodies the user's point of view about the data layers and the manipulation procedures that can be implemented on those layers as well as providing standard GIS capabilities.

The current capabilities of KBGIS include a KBGIS specific interface, which allows a user to query the system to find sets of spatial locations which the user described as well as to search for specific spatial objects that exist at a specifically predetermined location, a knowledge base which stores definitions and other useful information about objects in the system, and a learning procedure that is designed to reduce query search time. The system was developed on a VAX 11/750 using VMS and programmed in Common Lisp, C, and Pascal. Ongoing research at UCSB includes the improvement in file structures and indexing methods, integration of remotely sensed data, incorporation of useful image processing techniques and the extension of methodologies to improve vector data input. [ALB88]

National Space Science Data Center (NSSDC) KBGIS Development

KBGIS maintains each data layer in an indexed quad tree, and has several spatial operators to access and combine layers into a new layer. Knowledge about each layer is stored in a frame-based language, so that queries about the entire layer can be answered quickly. The quad tree allows a natural hierarchical pyramid of resolution that can be exploited during spatial searching to quickly eliminate quads that do not have the desired objects. As more is learned about a layer, that information can be added to the frame describing that layer, thus making the system smarter with use. This capability is a significant advancement over traditional GIS's. In addition to the spatial data, KBGIS has a frame-based layer of objects either extracted from the data, or inserted into the object knowledge base by the user. The objects may have a location attribute that points to where the object was discovered or derived, or they may be abstract types that help form hierarchies of the objects. Objects are represented in the Spatial Object Language (SOL), which allows complex, nested expressions of ANDS, ORS, NOTS, and other operators to compose objects of multiple parts. Also, to allow new spatial operators to be added easily to the system, KBGIS has a function knowledge base that allows users to add new operators to the query language. The operators can access and manipulate the data layers to perform discipline specific operations. [SMI87]

Through an informal memorandum of understanding with the USGS, the National Space Science Data Center (NSSDC) is currently developing a UNIX-based version of KBGIS implemented on a Sun 3/260 workstation. In addition, we are extending KBGIS to interconnect with existing NASA systems, such as imagery from the Land Analysis System (LAS), atmospheric data in Common Data Format (CDF), and visualization of complex data with the NSSDC Graphics System (NGS). These capabilities allow the utility of easily transferring data in and out of KBGIS. Figure 2 illustrates the specific work described in the next section.

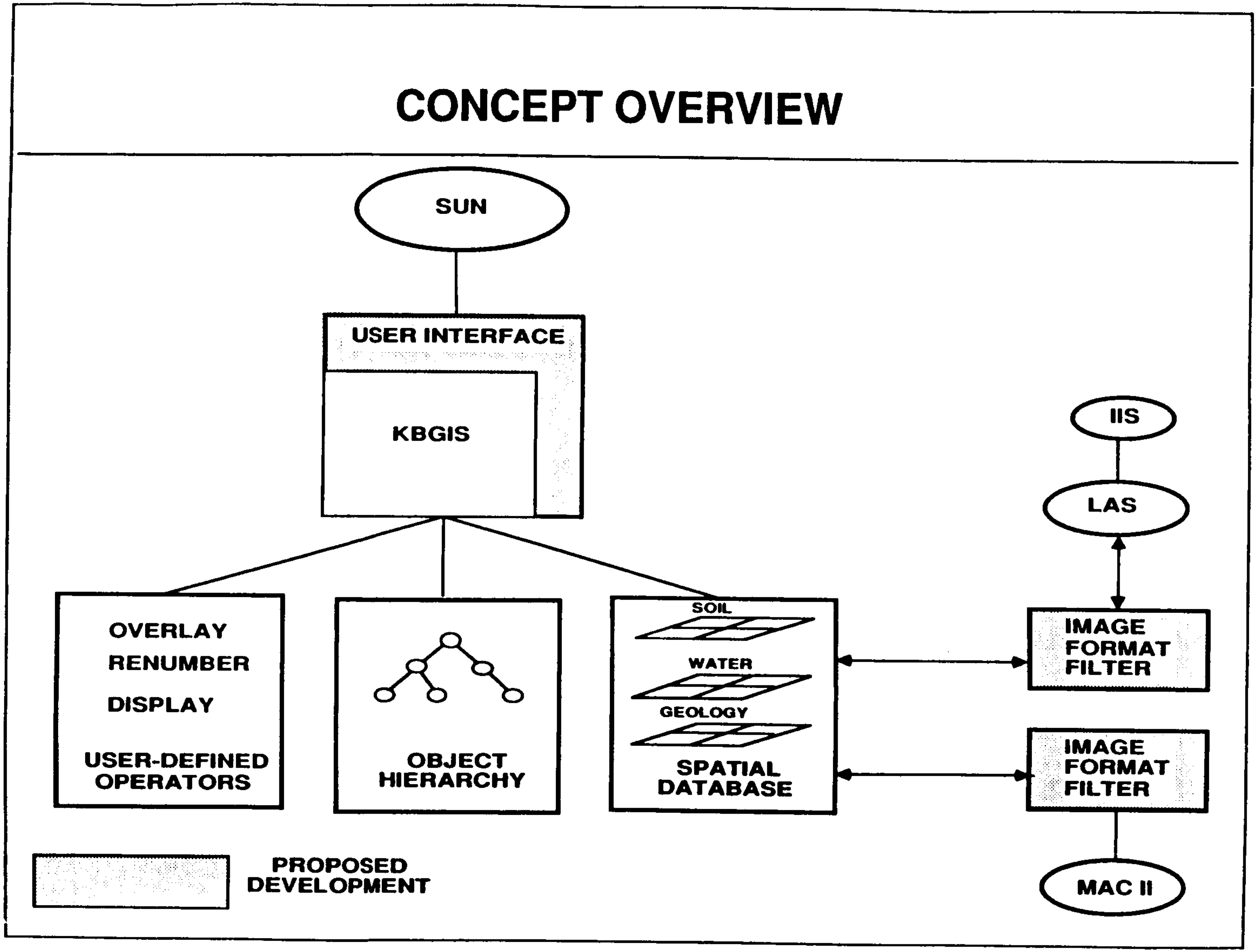
CONCEPT OVERVIEW
SUN
USER INTERFACE
KBGIS
OVERLAY
RENUMBER
DISPLAY
USER-DEFINED OPERATORS
OBJECT HIERARCHY
SOIL
WATER
GEOLOGY
SPATIAL DATABASE
IIS
LAS
IMAGE FORMAT FILTER
IMAGE FORMAT FILTER
MAC II
PROPOSED DEVELOPMENT

Figure 2.

Current Status

With the recent developments in computer technology, it is becoming unreasonable to demand that an application user be required to use a proprietary software system on a specific hardware platform. The same rules apply to GIS software. The evolving software standards, which have recently emerged (e.g., UNIX, X-Windows, PHIGS), allow software developers to design applications that can be transported to various hardware environments much more easily. The eventual goal of these software standards is to create applications that not only run on a variety of equipment, but also make the particular hardware environment transparent to the the user.

The immediate goal of the research on KBGIS at the NSSDC is to transform KBGIS into a portable system. As previously mentioned, the initial version of KBGIS was developed on a VAX. All of the low level input/output routines, which are required for all primitive object searches of the spatial location data base, are written with system specific (RMS) I/O functions. These system specific functions not only limit the software to be tied to the VAX, but force the data that has been stored using these functions to reside on the VAX as well. The first step of this transformation has been to rewrite the spatial location data base I/O routines to use a more standard UNIX I/O library. With the continuing increase in both CPU speeds and disk accessing speeds, there should no longer be a need to use system specific I/O routines. The indexing strategies that are offered with the VAX I/O routines have been rewritten with routines from the UNIX I/O library and provide the same functionality although at somewhat slower speeds. However, once the data exists in a format created by the UNIX I/O functions, it can be easily transported to other systems either directly, if both systems share the same file structure, or by transferring the data via some remote connection routines such as Remote Procedure Calls (RPC) and eXternal Data Representation (XDR). RPC and XDR are two examples of tools that allows software running on two different hardware environments to communicate and transfer data across a network. [SUN86]

The original KBGIS system only supported one format, the Earth Resource Data Analysis System (ERDAS) for import and export into the system. To promote greater use in the NASA community, several additional formats widely used on various NASA projects have been added to the system. The Land Analysis System (LAS) is an image processing system developed by NASA. LAS provides the capability to import, analyze and manipulate images. One of the main features of LAS that directly relates to KBGIS is the ability to register geometrically one image to some map (or ground) location or multiple images from various instruments to each other. With the ability to directly import LAS images into KBGIS, a vast amount of spatially distributed data becomes available to create new data layers that can be immediately preprocessed and analyzed by KBGIS. [NAS87]

The Common Data Format (CDF) is another format gaining support with increased use in the NASA community as well as other agencies. CDF is a software package written at NSSDC which consists of a library of routines that form a data-independent abstraction for multidimensional data structures. [TRE87] By adding the ability to export query results from KBGIS into CDF format, many of the applications written using CDF have also become available. One such application is the NSSDC Graphics System (NGS). NGS is another application developed at NSSDC to visualize scientific data. Many of the functions of NGS have been developed to support the visualizing of geographic data. For example, geographically stored data is ingested into NGS from CDF format and stored in an "oct-tree" data structure for quick retrieval of a particular geographic region. [SAM84] NGS also supports four-dimensional plot types with the use of a color lookup table and a mapping capability that supports over 20 general projections.

Although much research has been directed towards object hierarchy and spatial search, [SMI86] minimal effort has gone into developing the user interface, making it difficult to use the original system. For instance, the interface to KBGIS did not have knowledge of the domain's high-level concepts, but instead forced the user to render his query in terms of the low-level ordering of the spatial object knowledge base. With the recent emergence of the X-Window System and, more specifically, the X-toolkit (Xtk) as a standard library for window management, creating user interfaces which can be ported to various hardware platforms has become increasingly easier. With the use of the standard X protocol, applications can retain the same look and feel across different environments. As other layers of user interface tools become available, such as Open Look, development of these types of applications will become even simpler. [SCH86]

With the use of Xtk, objects known as "widgets" are used to create various user interface elements such as menus, forms, scroll bars and item selection buttons. An additional layer, using this core-widget set, is built onto KBGIS to provide a user-friendly interface into the system. A user is able to form spatial queries without any knowledge of the spatial object language (SOL), which KBGIS uses to represent objects. The system interacts with the user in terms of the domain knowledge which the system uses to directly access the spatial object knowledge base. The new software layer then transforms the user's input into a standard SOL query which then directs the system's search. As the user interacts with the system, the spatial knowledge base is dynamically maintained so that the system retains the modified or newly created objects defined by the user. This removes the burden from the user of having to render the queries in terms of pixel properties (e.g. agriculture, residential) or relational properties (e.g. distance), and permits him to interact with KBGIS at a higher level. This increases the ease of using the system, and decreases the amount of time required for learning KBGIS.

Future Directions

There are several areas of research that could be undertaken to enhance KBGIS.

A) A natural language front end could be added to KBGIS to give a user a more descriptive way of querying the spatial object knowledge base.

B) The current learning system in KBGIS could be enhanced by adding learning by example thus making it more autonomous.

C) The display graphics of the current system is limited. To bring KBGIS up to the functionality of other GIS products on the market, several display capabilities could be added.

1) Add cartographic capabilities for base maps and hard copies
2) Add ability for image overlays

D) Enhance KBGIS with a more intelligent and robust inference engine. This will be required as the system evolves and becomes a more general purpose system.

Summary and Conclusions

Work at the USGS, UCSB and NASA continues to enhance the functionality and capability of KBGIS. Currently the system can respond to various types of spatial queries of large complex, heterogeneous data in a multilayered environment. Work at UCSB is focused on increasing the efficiency of spatial search using heuristics as well as on a more friendly interface. Our work at Goddard is primarily focused on providing a methodology for importing external data sets, primarily enhanced imagery, in an easy fashion as well as concentrating on a standardized, generic graphics interface on a Sun workstation. Once this work is completed, we will evaluate the feasibility of linking and customizing a natural language query processor [NLQP] and a high level expert system that will deal with the pragmatic translation between the NLQP and KBGIS. We will also determine what role artificial neural networks can play in automatically characterizing external data into a form that can be readily utilized by KBGIS.

Acknowledgement

The authors would like to thank Scott Wattawa of the National Space Science Data Center for his thoughts, ideas and expertise that are presented in the concept formulation and interface design section of this paper.

References

[ALB88] Albert, T. M., "Knowledge-Based Geographic Information Systems (KBGIS) New Analytic and Data Management Tools", Journal of Mathematical Geology, 1988.

[CAM81] Campbell, W. J., "Geographic Information Systems - A Training Module", Chapter 7 in N. M. Short., Landsat Tutorial Workbook, NASA/GSFC, May 1981.

[NAS87] NASA, "LAS User's Manual - Version 4.0", GSFC/NASA, September 1987.

[SAM84] Samet, H., "The Quadtree and Related Hierarchical Data Structures", ACM Computing Surveys, 16, 2, 187-260, June 1984.

[SCH86] Scheifler, R., and J. Gettys, "The X Window System", ACM Transactions on Graphics, 5, 2, 79-109, April 1986.

[SMI86] Smith, T., D. Peuquet, S. Menon and P. Agarwal, "KBGIS-II: A knowledge based geographic information system.", Technical Report TRCS 86-13, Department of Computer Science, University of California, Santa Barbara, May 1986.

[SMI87] Smith, T.R., S. Menon, J. Star and J. Estes, "Requirements and Principles for the Implementation and Construction of Large-Scale Geographic Information Systems", Int. J. Geographical Information Systems, 1987. Vol 1. No. 1, 13-31.

[SUN86] Sun Microsystems, "Networking on the Sun Workstation", Sun Microsystems, Mountain View, Calif., 1986.

[TRE87] Treinish, L., and M. Gough, "A Software Package for the Data-Independent Storage of Multi-Dimensional Data", Eos Transactions American Geophysical Union, 68, 633-635, 1987.