THE DESIGN OF A GEOGRAPHIC INFORMATION SYSTEM UTILIZING THE SYSTEMS ENGINEERING APPROACH

by

Kenneth J. Cavanaugh Jr.

Project and Report submitted to the faculty of
Virginia Polytechnic Institute and State University in partial fulfillment of
requirements for the degree of

MASTER OF SCIENCE

in

Systems Engineering

APPROVED:

Benjamin S. Blanchard, Chairperson

L. Kenneth Harmon

Holly Butler

September 9, 1994

Blacksburg, Virginia

THE DESIGN OF A GEOGRAPHIC INFORMATION SYSTEM UTILIZING THE SYSTEMS ENGINEERING APPROACH

by

Kenneth J. Cavanaugh Jr.

Benjamin S. Blanchard, Chairperson

Systems Engineering Department

(ABSTRACT)

A geographic information system, GIS, is a computerized mapping system which assists in organizing, analyzing, and visualizing spatial data sets. Over the last decade, many local and state agencies as well as many commercial organizations have implemented numerous GIS applications to raise their staffs' productivity when working with spatial data. A GIS provides organizations the ability to consolidate data from numerous hardcopy maps into a single spatial database which may be accessed by different internal organizations for data analysis and output. A local government may use a GIS to manage tax parcels and urban development, to identify areas with reoccurring criminal activities, or to track the location of assets such as manholes, water and sewer lines, and roads. In a majority of existing GISs, system operators, or analysts, identify the locations of individual map features by overlaying these features against a base map. The base map may be a raster image such as a scanned paper map or an overhead photo, or may be set of vector layers which consist of digitized roads and buildings from scanned maps or overhead photographs.

For local governments, the extent of their area of interest is small in contrast to the extent of the Earth. With small areas of extent, local governments can dedicate a majority of their GIS development budget to digitize accurate base maps. For organizations interested in tracking features across the Earth, the costs of digitizing features with a similar degree of accuracy is prohibitive. Based on the availability of base map data sources and the operational requirements, system designers of a global GIS must select the best spatial format (raster or vector) in which to represent and store spatial base map data. The selection of the format for the base maps should take into account the effects that the format will have on the hardware configuration, the display and network performance, and the maintenance and operational costs.

This project presents the conceptual and preliminary designs for a global geographic information system utilizing the systems engineering approach. The system addresses the need to reduce the time analysts require to complete an analysis task by reducing the time analysts spend researching, collecting, and assembling spatial data and by enhancing their ability to enter and analyze spatial data sets. The system provides for a consolidated spatial database based upon a standard set of base maps. The design focuses on the implementation of the base map database which will contain map sets at various map scales. The project evaluates the financial and performance costs associated with implementing these map sets as either a raster or a vector base map database. Based on the operational requirements for a petroleum consortium, recommendations are made to utilize an existing commercial vector product for the 1:1000000 scale base maps and to create a raster map library for the 1:250000 and 1:50000 scale base maps. A preliminary hardware configuration is presented which reflects configuration recommendations for utilizing a raster base map database.

ACKNOWLEDGEMENTS

I would like to thank my committee members: Dean Blanchard and Ken Harmon for their time, flexibility, and individual contributions with respect to the application of the system engineering process; Holly Butler for her time, ideas, and detailed critique of my project throughout its development. Holly's experience proved to be a valuable asset in the formulation of my evaluations and system specifications.

I would like to thank Martin Marietta and its Advanced Education Office for its support of continuing education. In particular, I would like to thank Christina Waldorfer for the outstanding job she has done in reorganizing the advanced courses program and for her constant attention toward the personal and professional development of the program's participants.

And most of all, I wish the thank my wife, Donna, for her love, patience, and support throughout this project and the entire Advanced Course program.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SECTION 1

INTRODUCTION

A geographic information system, GIS, provides an organization with the ability to consolidate spatial data into a single spatial database. The spatial database provides a resource to various internal organizations in the support of applications such as property/asset management, decision support, proximity analysis, and modeling. For systems which track physical features, a GIS maintains spatial information with regard to a map feature's physical specifications and location as a single vector feature (a collection of connected coordinate pairs). Each vector feature is assigned a unique identification number which is used to linked to external databases, text files, and photographs.

The most common methods of entering vector map features into a GIS are through coordinate entry using a Global Geopositioning Satellite System (GPSS), digitization using a digitizing table, or manual placement on the screen using digital base map data. The manual placement of features is a convenient method for analysts to enter their features of interest. The digital base maps used to assist in the manual placement of features may be represented and stored in either a raster or vector format.

Raster base maps consist of overhead imagery and scanned paper maps which are displayed as color images on the screen. These images are scaled and rotated to best fit the coordinate system (eg. longitude and latitude coordinates) of the display window. Analysts create their features of interest in vector format using the raster base map as a background reference. Figure 1 shows pipelines (yellow lines) and valves (blue triangles) as the analyst's features of interest which were entered against the raster base map, a satellite image, of Kuwait City.

Figure 1 Raster Base Map with Vector Features of Interest

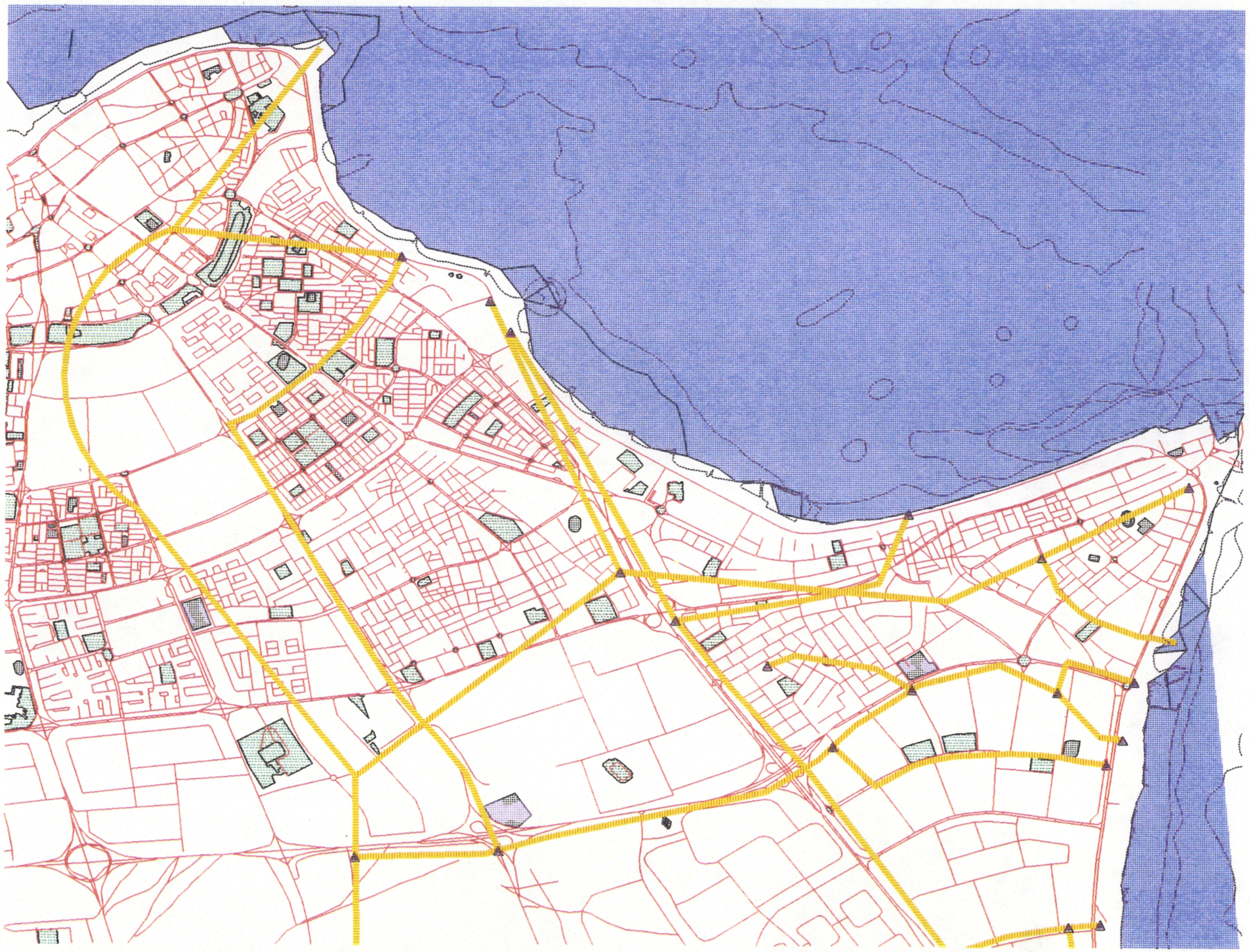


Figure 2 Vector Base Map with Vector Features of Interest

Vector base maps are digitized representations of the overhead imagery and paper maps. Individual map features are organized into themes and represented as points, lines, and polygons (regions of areas). In this case, analysts create their features of interest in vector format using the vector base maps as a background reference. Figure 2 shows an analyst's features of interest (pipelines and valves) which were entered against a Kuwait City vector base map consisting of roads (red lines), water (blue polygons) and vegetation (green polygons).

In this project, the base maps represent map features which are secondary to an organization's mission to track individual features. The base maps are used only as references in support of analysts' efforts to add their features of interest which assist in their analysis tasks. The base maps which present the most value to the analysts are the ones providing the greatest level of detail and accuracy.

The base maps provide a critical role in the creation and maintenance of various map features of interest. For example, the accuracy of features entered manually against the base maps can only be as accurate as the base map themselves. In addition, the base maps should contain a sufficient number of features (cities, roads, airports, etc.) which provide the analyst with a good reference as to where new features should be placed. When working with a small area or extent of the earth, it is feasible to build base maps which have a wealth of feature information and present a high degree of accuracy (+/- 6 inches). When working with an extent which encompasses the entire world, the costs associated with building a base map database with the same amount of feature information and accuracy become exorbitant.

Organizations interested in creating and tracking features across the entire globe face difficult decisions on how to best represent their base map database in order to provide their analysts with a common reference having sufficient accuracy at a reasonable cost. The decision to represent a set of maps in either a raster or vector format affects the system's development costs (architecture), productivity costs (performance), and maintenance costs (data manipulation and updates).

Why Perform the Study?

In order for a geographic information system to have any value, the system must have data sets which provide sufficient accuracy and feature representation to address an organization's data entry and analysis needs. Organizations which maintain the locations of physical features which reside across the Earth rely on numerous analysts to locate, shape and attribute these features of interest. These analysts are located in different facilities creating features of interest which reside in similar extents of the world. With vast amounts of spatial data residing on thousands of maps stored in hundreds of map cabinets across several facilities, organizations need a system which will provide their analysts with the ability to enter, update, and share their own data as well as the ability to access and analyze data from other facilities. The system would provide configuration control of individual features and allow analysts fast access to data based on boolean and spatial queries. Since most of these analysts enter new features against a base map reference, it is necessary that the organization provides a standard set of base maps from which all analysts add and position new features. By using a standard set of base maps, new features from several facilities can be consolidated into a single database with little or no manipulation and with no question with respect to the quality of the data source from which the features are created. An organization's decision on how to represent and store the base maps, in terms of raster or vector format, should be based not only on the costs associated with developing the base map database but also the costs associated with maintaining and displaying the base map data.

Project Scope and Methodology

The project begins with the need of a consortium of petroleum companies to increase their analysts' productivity. It is assumed that the consortium of has determined that a geographic information system will address its needs and would like recommendations with respect to how to best implement (raster or vector) its base map database given the operational requirements and cost factors. Following the definition of need, a needs analysis is performed for the entire GIS in Section I. The results of various feasibility studies, that are assumed to have been performed by the consortium, are also presented in Section I. Using the feasibility studies and needs analysis, the system's operational requirements and maintenance concepts are formulated and presented in Section II. It is assumed that these early system specifications are approved by the Conceptual Design Review Board. With the board's approval, the preliminary design begins. Section III presents three levels of functional diagrams for the Operations function of the system. A high-level allocation of requirements is also presented for functional units that are assumed to have been identified through the completion of the functional diagrams. Following the allocation of requirements, the project begins the first iteration of optimizing the preliminary system specifications. The focus of the optimization is how to implement the base map database and how the base map implementation will affect the preliminary system configuration.

Provided with three global map sets at various scales, the project evaluates how the two formats affect the system's network and drawing performance, the system configuration, the presentation and usage of data, and the life-cycle costs. Based on the operational requirements, recommendations are made with respect to the implementation of the base map database for each set of maps and the implementation of the system's architecture. While the project ends with these recommendations, it is assumed that the optimization process would continue to further define the complete preliminary system specification.

Systems Engineering Process

The design of this geographic information system utilizes the system engineering process as presented by Dr. Benjamin S. Blanchard and Dr. Wolter J. Fabrycky. By utilizing the systems engineering approach, the operational needs are transformed into a description of system performance parameters. The needs are later transformed into a preferred system configuration through the use of an iterative process of functional analysis, synthesis, optimization, definition, design, test, and evaluation. By using the system engineering process, performance, producibility, reliability, maintainability, manageability, and supportability, are integrated into the overall engineering effort. (1)

The system engineering process, Figure 3, is a series of steps which starts with the identification of user's needs and ends with the development of an optimized and effective system. The process consists of a number of phases which includes the definition of need, the conceptual design, the preliminary design, the detail design, production, operations and maintenance, and retirement.

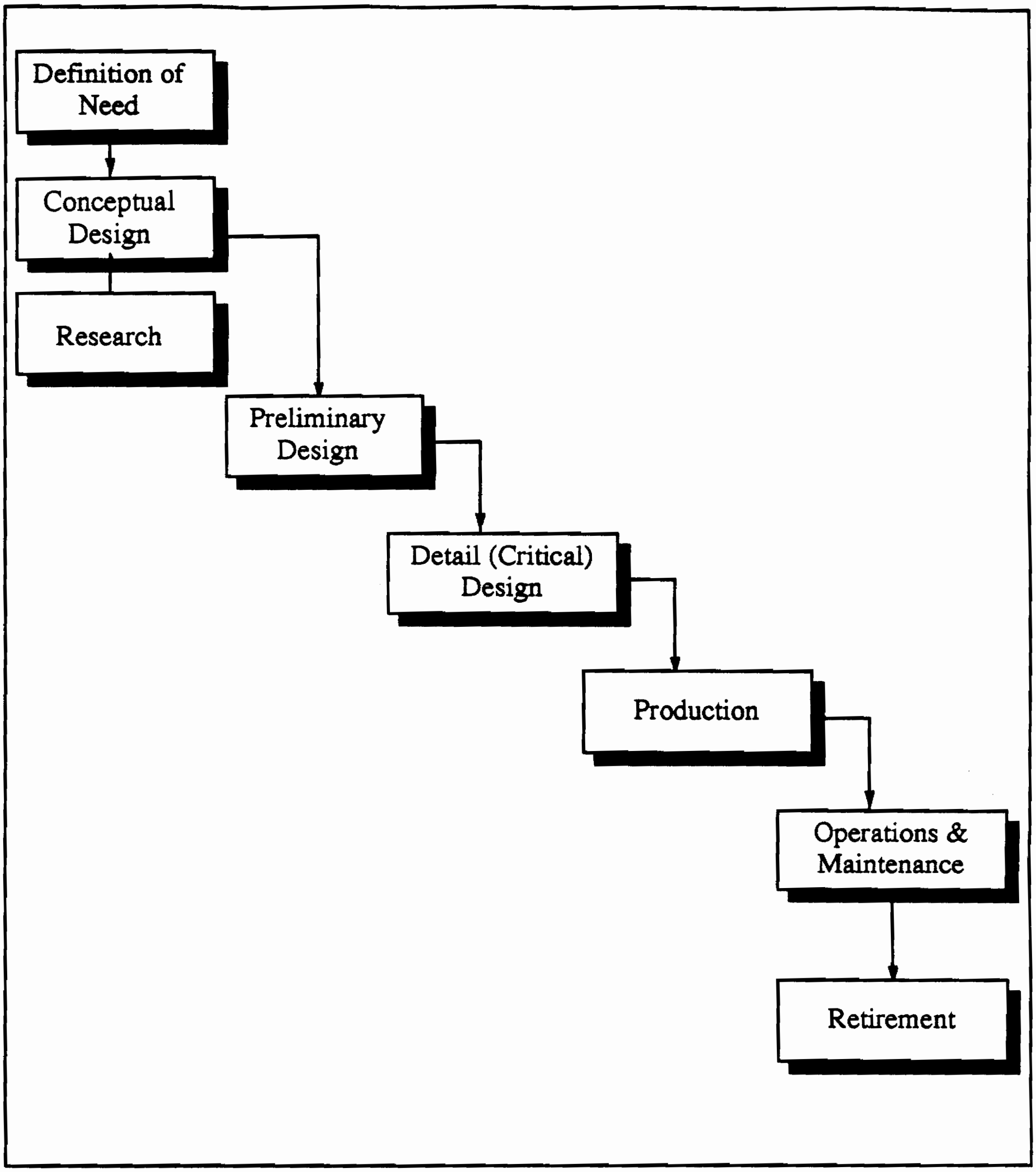


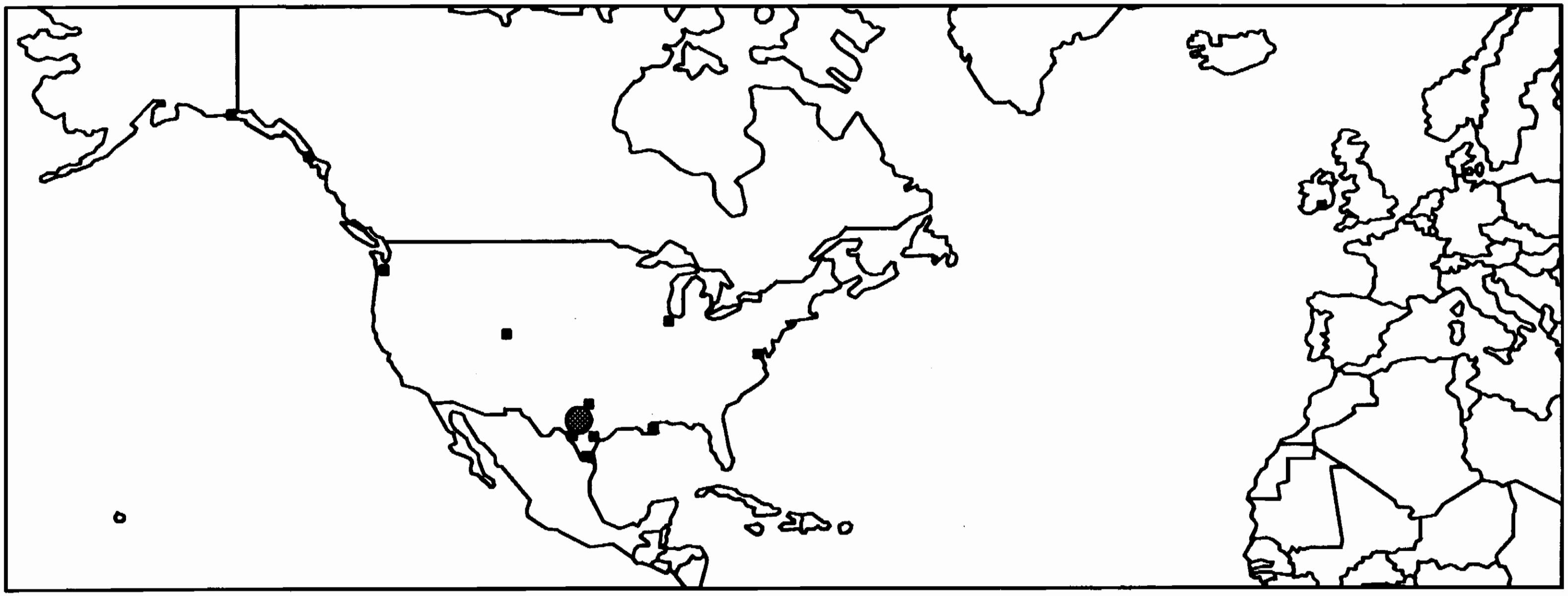
Figure 3 The System Engineering Process

Definition of Need

Each system begins with the identification of a need. This need is usually the result of some deficiency or problem that requires resolution or implies the desire for an added capability. In identifying the need, the systems life–cycle approach begins. The objective in defining the need is to address the fact that the need actually exists and is not just a perceived need. (2) This section describes a petroleum consortium's need to implement a geographic information system in order to assist the consortium, as well as its members, with their mission of tracking the locations of pipelines located around the world. The system addresses the following needs:

- Reduce man-hours spent searching for spatial and attribute data sources from 80 hours to 24 hours per analysis task.
- Reduce man-hours spent producing hardcopy maps from 40 hours to 8 hours per analysis.
- Increase analytical capabilities through overlay and query analysis of pipeline map features (pipelines, pump stations, access hatches, valves).
- Consolidate pipeline map features into a single spatial database.
- Producing a standardized set of base maps, within provided funding, on which all pipeline data is entered to ensure accurate feature positioning.

The petroleum consortium was recently established by ten petroleum companies, Figure 4, in response to environmental regulations which called for a single organization to track information and locations on pipelines which transport petroleum products for the individual companies. The consortium is responsible for obtaining pipeline data from the companies' research departments and consolidating the data into a single pipeline database from which the consortium's analysts fulfill analysis tasks assigned by federal and international agencies.



Central Facility:

Consortium Headquarters (Austin, Texas)

Remote Facilities: (Small)

Eagle River, Alaska
Ankorage, Alaska
Seattle, Washington
Aurora, Colorado
San Antonio, Texas

Remote Facilities: (Large)

Houston, Texas
Dallas, Texas
Chicago, Illinois
Dover, England
Fairfax, Virginia

Figure 4 Central and Remote Facility Locations

The petroleum companies' research departments are already operating at their top capacitities in terms of the number of analysis tasks handled a year. The companies are concerned that they will not be able to satisfy the analysis tasks requested by the consortium for their entire pipeline systems in addition to their own internal analysis tasking. Future staffing restrictions limit the number of analysts within each company's research department. The departments can only expect staff increases of one or two people over the next five years; so the companies are looking for a way to increase the number of analysis tasks handled by an analyst in order to address the consortium's data needs as well as the companies' growing analysis needs.

The Petroleum Companies

Each company contains a pipeline research department whose purpose is to perform analysis tasks in support of other divisions' operations such as maintenance planning, development, exploration services and safety support. These research departments store information associated with previous analysis tasks in hardcopy products such as presentation charts, paper maps, and loose-leaf binders. The research departments rely on this data as well as data it can obtain from other internal divisions to fulfill its analysis tasks. Analysis tasks are assigned by each company's Analysis Review Board (ARB) which receives the analysis requests from various internal divisions. The ARB validates the need and priority for the analysis and assigns the research department with the analysis task and its requirements. The ARB receives more analysis task requests than can be handled by the analysts in a single year. On average, companies can only fulfill 80% of their analysis task requests.

Upon receiving an analysis task, an analyst spends approximately 40 hours gathering data from within the research department's map cabinets, binders, and databases. For approximately 50% of the tasks, an analyst requires data which does not exist within the research department. An additional 40 hours is required to obtain and analyze data outside of the research department. Once the analysis is complete, an analyst spends 40 hours assembling a presentation showing the results of the research. The presentation usually contains two map products which are produced using color tape, a label machine, and laminated paper maps. Following the presentation, the maps and research data are stored within the research department to assist in future analysis tasks.

Five of these companies have research departments staffed with two analysts. Each of these companies have pipeline networks consisting of approximately 2000 pipelines in only 1-2 continents. The two analysts are responsible for all analysis tasks pertaining to these pipelines. Each research department maintains two map cabinets holding approximately 1000 different maps from previous analysis work. A research department with two analysts can handle approximately 42 analyst tasks in a year.

The other five companies have research departments divided into two divisions. Each division has six analysts. Research is divided between the divisons based upon whether the analysis requirements lie in the northern or southern hemisphere of the Earth. Each of these companies have pipeline networks which consist of approximately 10,000 pipelines. Their pipelines exists between the latitudes of 80N and 60S. These research departments maintain 5,000 maps from previous analysis work in eight map cabinets. A research department consisting of 12 analysts can handle approximately 250 analysis tasks in a year.

With the creation of the consortium, all ten of these companies' research departments are expected to perform not only their current analysis tasks but they are also required to assemble and maintain information describing their entire pipeline network for inclusion into the consortium's pipeline database.

The Data

All analysis tasks are related to locating selected company pipelines within a specified geographic region. The spatial data of interest to analysts includes the location and shape of a pipeline; the start and end of a pipeline; the locations of pump stations and access hatches along a pipeline; and the proximity of a pipeline to base map features such as roads, rivers, and cities. This data would be most accurately represented as topologically defined vector map features such as points, lines, and polygons.

The analysts also usually require attribute information on a pipeline's technical specifications, installation and maintenance history, and manufacturer. A limited presentation of this information may exist in an existing map product's legend or in related documentation. A pipeline's detailed specifications are maintained in free-formatted project documentation, schematics, and engineering drawings. This data is maintained in separate binders and is difficult to find unless a reference number has been assigned to a map product containing the pipelines of interest. Text search/retrieve capabilities and database links to individual map products or pipeline features would be very beneficial in the gathering of analysis data. Image scanning and optical character recognition capabilities would provide the analysts with on-line access to these free-formatted data sources.

The analysts use data gathered from their research to plot the location of pipelines of interest. These pipelines are positioned on a map using survey data such as measurements from a reference marker or coordinates from a geopositioning satellite (GPS) receiver. While GPS receivers provide highly accurate placements for some of the companies' pipelines, many of the pipelines are positions based on a measurement from the edge of a road between mile markers or cities which are present in a base map. The base maps are required by the analysts for locating and placing pipelines on the ground. Since heavy equipment must have access to an area in order to install a pipeline, most pipelines are installed along transportation structures such as roads and railroads. Transportation features as well as city locations usually provides sufficient locational information for the placement of pipelines. Proximity presentations are assembled with base map features showing environmentally sensitive areas, population areas, and petroleum industrial areas.

The research departments produce map products at various scales depending on the available base maps. Most analysis tasks and presentations results are located within a one or two countries; base maps with map scales of 1:1,000,000 to 1:50,000 are usually utilized for these presentations. The presentations highlight selected pipeline features of interest against a base map showing transportation features, rivers, country borders, and cities as locational references. The presentations contain map legends, borders, titles, and north pole directions.

The research departments maintain a majority of the data with regard to their companies' pipelines, but not all the data. Future analysis tasks to provide the consortium with their company's pipelines will require that the research departments obtain and maintain current maps describing their entire pipeline infrastructures.

The analyst's presentations also include user-defined map features which are not pipeline or base map features but are features that are of interest to the analysis being performed. Examples of these kinds of features include dirt access roads or mile markers. These user-defined features are important in the presentation of analysis results.

The Analysis Tasks

A petroleum company maintains a wealth of data, approximately 500 pages a pipeline, on the location and specification of its pipelines. Research on this data requires many man-hours since the data may reside in various maps, studies, surveys, and projects. Collecting information on a single site involves finding relevant maps, comparing the maps side to side, and extracting information for the manual creation of a new paper map. In addition, analysts have difficulties in finding supporting documentation defining map features. The following paragraph describes the tasks involved with providing the consortium with pipelines for a specific map extent using the current research methods.

Analysts track the location of their companies pipelines on individual paper maps of various scales and quality. In order to produce a map for the consortium which contains only pipelines for a particular map extent, an analyst must search through the research department's map cabinets and research the pipelines' specifications through various sets of binders and databases which are referenced by the pipeline features. Once the information has been collected, an analyst must procure a new base map on which the research is assembled. Two copies of the finished map product are made. One is added to the company's map cabinets; the other is sent to the consortium. Once the consortium receives the map and supporting documentation, a consortium analyst would replot the data from the submitted map onto their own map products. Using these methods, analysts spend more time locating and assembling data sources rather than analyzing them.

Internal analysis tasks, exclusive of the consortium's data needs, center around locating and identifying individual pipelines within a specific geographic region. The accuracy required in the presentation of this data does not have to be stored at accuracies required by the engineering and installation divisions. Location of pipelines within 200 feet is sufficient. Once the pipelines have been identified, the appropriate divisions responsible for the selected pipelines provide detail locations based on the installation engineering drawings.

Besides identifying the location of selected pipelines, additional analysis is performed to show the proximity of pipelines to map features such as national parks, wildlife refuges, areas of population, petroleum industrial areas. Presentations usually show a general proximity based on the placement of pipelines with respect to map features of interests such as national parks. This is usually accomplished with the overlay of transparent presentation slides against a base map for the region of interest.

In addition to the pipelines, analysis tasks include identifying the locations of pump stations, emergency shutoff valves, and pipeline maintenance access hatches. Analysts are primarily interested in the locations of these places along the pipeline; however, information concerning these points' operational status, name or designator, and address is valuable to certain analysis tasks.

The following bullet items present examples of analysis tasks assigned over the last year at a selected facility:

- Find all pipelines installed between 1972 and 1985 which have been in active service for at least three years.
- Find all pipelines manufactured by the ACME Pipe Company whose lot number is greater than 9995.
- Identify the locations of all physical structures within 1000 feet of the K9-Alaska pipeline between markers 456 and 470.
- Identify all petroleum storage facilities which have been built within 800 feet of a A5 type pipeline operating at a pressure threshold greater than 5% of the pipelines pressure safety limit.
- Identify the pipelines and their pump stations that are operating under the series G software upgrade.

The Petroleum Consortium

The consortium was recently established in response to environmental regulations which called for a single organization to track information and locations on pipelines which transport petroleum products for the individual companies. The consortium will act as the custodian for pipeline information pertaining to its company members. In addition to the tracking of pipelines, the consortium is funded to perform research projects that are of interest to all of the petroleum companies. Examples of this research include independent reviews of emergency response plans and environmental impact studies for various regions.

The consortium is responsible for obtaining pipeline data from the companies' research departments and consolidating the data into a single pipeline database from which the consortium's analysts fulfill analysis tasks assigned by federal and international agencies. These analysis tasks are similar to the tasks handled by the individual research departments with the exception that all consortium member pipelines are of interest rather than a single company's pipelines. Using current research methods, the consortium would be consolidating pipeline data sent from the companies' research departments consisting of various map products, pipeline documentation, and schematics. Each map sheet would provides its own story in terms of the amount and type of data as well as the level of detail and accuracy. This would require the consortium's analysts to manually replot a company's pipeline data and to resolve any accuracy and missing data issues with the respective company's research department.

The Pipeline Tracking System (PTS)

To reduce the amount of time an analyst spends performing research, collecting data, and producing maps and to facilitate the consolidation and analysis of pipeline data within the consortium's database, the consortium has proposed the development of a GIS called the Pipeline Tracking System, PTS. PTS will provide analysts with geoprocessing capabilities to create and manipulate pipeline features at their local terminal. These features may then be submitted to the database administrator for the remote system who then submits the pipeline features to the consortium for inclusion into the consortium's pipeline database. The PTS would increase an analyst's productivity by increasing the availability and access to corporate spatial data, reducing the time required to collect spatial data and produce map products, and enhancing his or her analytical ability through query and overlay analysis.

For a number of years, all of these companies have seen the need to automate their access to corporate spatial data. The larger companies, who have inventories of pipelines that exist all over the globe, have been actively investigating the need to enter and access the data using a spatial database. By consolidating their spatial data, their analysts will become more productive by performing research faster through boolean queries and by having access to related and up-to-date data from other internal divisions.

Efforts to build a consolidated spatial database require a database design which must be integrated with the consortium's pipeline database. In addition, consolidating data into a single coordinate system (geographic coordinates) requires background maps or a base map database when the latitude/longitude coordinates of the pipeline feature are not known. The base map database would consist of several thousand individual maps at various map scales. The cost of electronically storing all of these maps in a single database cannot by justified for use in only one company. Since all of the consortium members are interested in building a spatial database of pipeline features within their own companies based on a standardized base map database, the companies have sought to pool their resources together. The consortium has allocated 28 million dollars over the next seven years to fund a single system which will support the consortium's mission as well as their own needs to consolidate data.

In order to facilitate the consolidation and portability of data, the companies have tasked the consortium to build a base map database which stores base maps at three scales: 1:1,000,000, 1:250,000, and 1:50,000. The analysts require the following themes of data in their base maps to assist them in the pipeline data entry and analysis:

- country borders
- transportation features, including roads, railroads, airports (since many pipelines run parallel to transportation features)
- city locations, including city limit borders if available
- wildlife refuges, parks, and other environmentally sensitive areas
- drainage features, including streams, lakes, swamps
- geological features, including soil types, fault areas
- petroleum features, including industrial areas, storage facilities, wells

These themes of data provide the analysts with information concerning where to locate a pipeline given a lack of geographic coordinate data.

The consortium will receive $2,000,000/year for the next seven years for base map production. The base maps can be represented in either a raster (image) or vector (coordinate) format. The consortium would like recommendations on how to represent each map set based on the life-cycle costs, presentation, and performance characteristics associated with each format. Life-cycle costs and schedule have the greatest priority followed by performance characteristics, presentation, and disk storage.

The consortium is looking toward implementing a geographic information system with a central facility at its headquarters and remote facilities at the various member facilities, Figure 5. The central facility will be responsible for the production of base maps, the configuration control of all base map and submitted pipeline data, and analysis tasks assigned by international and federal agencies.

The data under configuration control, base maps and submitted pipeline data, will be kept on–line or near on–line (optical jukebox) for analysts to access from their local terminals. The base map database will consist of the vector data set ArcWorld (1:3000000), commercially available vector city and country maps, as well as data from the three map sets provided to the consortium. While the central facility and the large remote facilities require on–line access to the entire base map database, the small remote facilities will only require a third of all the base map data since their area of interest is limited to part of the Earth.

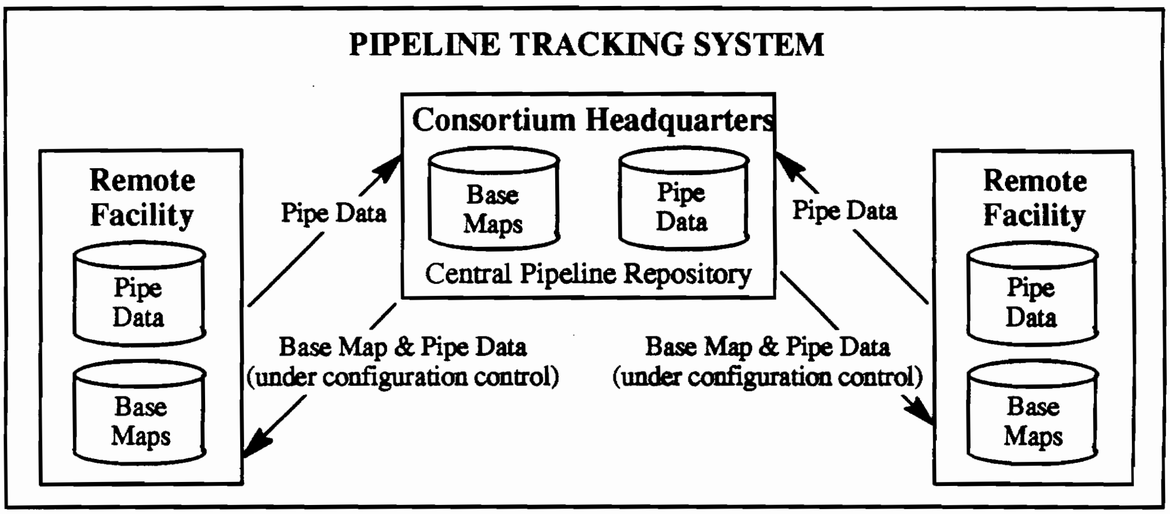


Figure 5 Pipeline Tracking System

The remote facilities have been divided into small and large remote facility classifications. Since each remote facility has existing, yet different, hardware, software, and network configurations, each company is responsible for the design, acquisition, and maintenance of their individual systems. The PTS design will propose standard configurations for both small and large remote facilities to assist the facilities in their design. The PTS design will present the required configuration for two remote facilities: San Antonio, TX (small facility) and Fairfax, VA (large facility). The other companies' remote system designs should reflect similar hardware performance and storage specifications as well as utilize all Commercial-Off-The-Shelf (COTS) software packages under the specified version and operating system. The San Antonio facility currently has a Novell network of 386 PC clients with MSOffice and Lotus Notes software packages. The Fairfax facility has a ethernet network of Sun Microsystem Sparc1s, IPCs, and 10s with Framemaker, CCMail, and Arc/Info GIS applications. Due to advancing hardware and software technologies as well as existing investments in the companies' system configurations, the PTS system should conform to industry open system standards and be designed for a scalable architecture to accomodate growth.

In order for the system to be an effective analyst tool, the performance of the system should not hinder the analyst's ability to be productive. The display and retrieval rates are deemed to be most critical in assisting the analyst in his or her tasks. Using the digitized and rasterized Defense Joint Operations Graphic (DJOG) for Kuwait, system performance rates are provided which have been judged by the analysts as acceptable performance specifications. The drawing rate for the culture points, transportation lines, and vegetation polygons should not exceed 27 seconds for vector and 35 seconds for raster. The retrieval rates for queries against a vector coverage which satisfies a three operator boolean query will exceed a rate of 4 MB/sec. The initialization time will not exceed 45 second. Vector map products will be produced in less than 300 seconds on color and black and white printers. Raster map products will be produced at rates greater than 0.009 MB/sec. Scanning rates for letter-size pages will exceed 8 pages per minute; scanning times for an E-size map will be less than 15 minutes.

Using the present method to produce a new map product over forty hours, analysts spend most of their time replotting and transfering data from one map product to another. The PTS should emphasize the portability of data among the various facilities through on-line file transfers and magnetic and optical media (8mm tapes and Read/Write Optical Disks). Portable data provides for increased data sharing and reduced data replotting.

With the population of data occurring at various remote facilities, the system should provide configuration control of the pipeline data at two distinct levels. The first level of control is within the various remote facilities. At this level, pipeline features, entered by the analysts, and their associated attributes are submitted to the company's database administrator, DBA, for inclusion into the company's pipeline database. Following internal configuration control procedures, the company DBA places the analyst's submitted map features into configuration control. If the submitted map features impact the consortium's pipeline database, these pipeline map features are forwarded to the consortium's headquarters for inclusion into its consolidated pipeline database in accordance with established configuration management procedures. Inclusion of company submitted pipeline map features by the consortium DBA is the second level of control. The remote facilities may order copies of the consolidated pipeline database for their review and analysis. The base map data is controlled at one level, the consortium's headquarters. Any discrepancies in the base maps must be reported to the consortium's DBA. The consortium investigates base map discrepancies and updates the base maps when necessary.

The consortium's pipeline database consists of pipeline features that are added to the display by analysts at the various remote facilities. Using base maps provided by the consortium, the analysts use a graphical user interface which provides them easy access to various COTS products as well as to the GIS product's display, editing, and analysis capabilities. Once entered, new features may be submitted for inclusion into a pipeline database which is under configuration control. Submitted data from remote facilities is received by the central facility in a digital format which can be directly entered into the consortium's centralized database.

In addition to entering pipeline data for the consortium, some of the individual petroleum companies plan to use the system to track additional features of interests so as to build their own consolidated databases. This additional development is performed on an individual basis by each company.

Figure 4 shows the location of facilities in which PTS will be deployed. The central facility resides at the consortium's headquarters in Austin, Texas. Headquarters is responsible for the creation and distribution of base map data. The headquarters facility will also provide configuration control for both the base maps and any pipeline data submitted by remote facilities. The remote facilities have been organized into two categories. The small remote facilities are projected to have two analysts. The small remotes are located in Alaska, Washington, Denver, and Texas. The large remote facilities plan to support multiple divisions at different locations each with a varying amount of analysts (2 to 6) in a division.

Feasibility Studies

Commercial Off The Shelf (COTS) Products

Based on the evaluation of geographic information systems implemented at various government offices and engineering firms, the consortium has concluded the system will focus on implementing system functions through the integration of available commercial products which may be customized by application. By using COTS software and hardware, the consortium will achieve its operational goals with a minimum investment when contrasted to the funds required to develop a geographic information system from scratch. In addition, maximizing the use of COTS products reduces the dependency on custom applications which will reduce future maintenance and support costs.

Open System/Scalable Architecture

The consortium's members have a substantial investment in existing networks and systems. The majority of the systems utilized PC and Unix platforms operating on Ethernet networks (TCP/IP protocol). With rapid changes in hardware technologies, the design shall focus on having an open system architecture. Conforming to open system standards, the consortium hopes to reduce duplications of systems, increase resource reuse, and provide network scalability and flexibility.

ARC/INFO GIS Software

Based on the evaluation of industry's largest GIS software providers, the consortium suggests the use of Environmental Systems Research Institute's, ESRI, ARC/Info software for the creation, manipulation, display, and storage of spatial information. The consortium favors ARC/Info based on the following:

- ESRI's financial strength and large market presence.
- Majority of consortium members utilize ARC/Info software.
- ARC/Info's geoprocessing tools provide for the basic geoprocessing capabilities believed to be required by the proposed system.
- Provides the capability to build customized applications through the use of the Arc Macro Language.
- Provides the capability to interface with relational database management systems.
- Provides a large number of spatial data conversion tools.

Relation Database Management System (RDBMS)

Based on the evaluation of ESRI's Info database (limited capabilities) and the large number of existing, corporate databases which reside in relational database management systems, the consortium has concluded that the system shall provide access to external relation databases for the storage of non-spatial attributes.

SECTION 2

CONCEPTUAL DESIGN

The conceptual design takes data from the needs analysis, research, and feasibility studies and performs activities which produce an early specification of the system. These activities are shown in Figure 6. The needs analysis and the results for the feasibility studies are shown in Section 1, Introduction. The following section presents the operational requirements for the Pipeline Tracking System.

Primary Mission Statement

The Pipeline Tracking System (PTS), Figure 7, will provide a centralized repository of spatial and attribute data for pipelines owned/operated by consortium members. The system will provide functionality allowing members to enter and store data and provide query and display functions to locate and isolate individual features. The system will track the physical location of pipelines as well as the pipeline's specifications which are stored as database attributes, text and image files.

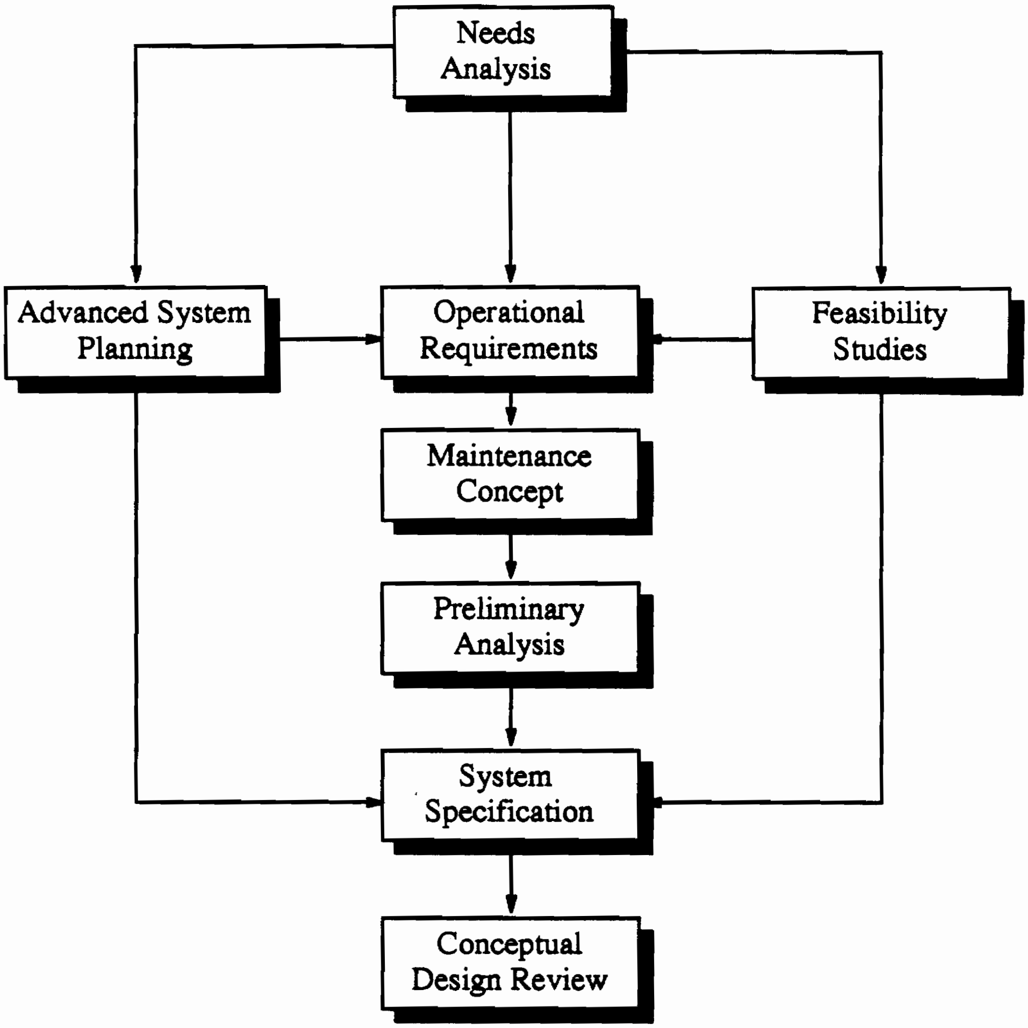


Figure 6 Conceptual Design Activities

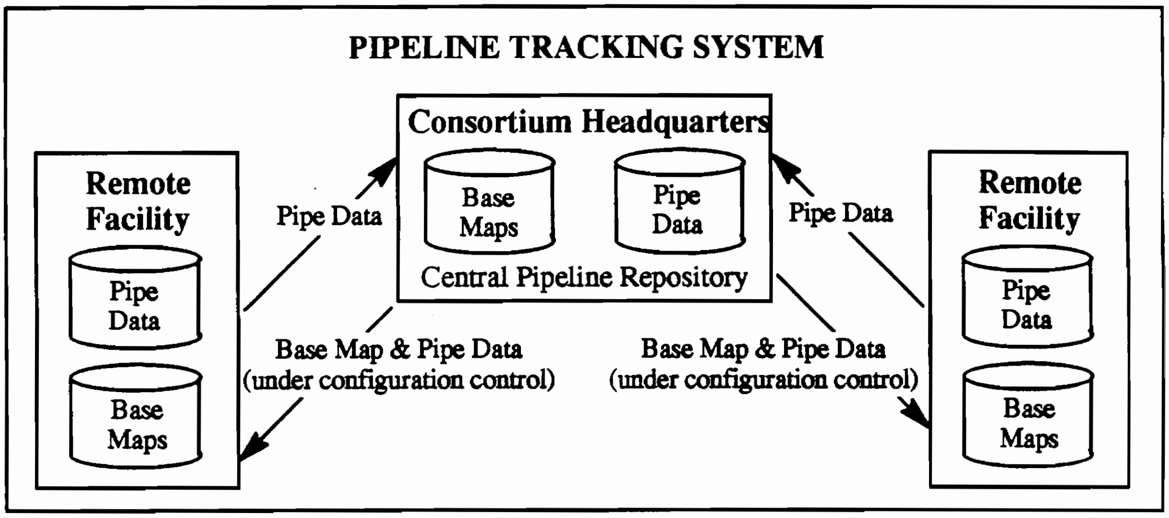


Figure 7 Pipeline Tracking System

Secondary Mission Statements

The system will provide analysts with access to a standardized set of base maps, having map scales of 1:1000000, 1:250000, and 1:50000, that are available from the consortium.

The system will provide a wide-range of data entry and geoprocessing tools allowing analysts to organize, add, manipulate, and store new spatial features.

The system will provide the capability to scan and digitize hardcopy maps for use as base maps.

The system will provide a query mechanism to allow analysts to locate map features based on a boolean query against the feature's attribute data fields.

The system will provide a composition mode allowing analysts to create cartographic compositions consisting of raster and vector data, annotation, simple graphic shapes, and text files for presentation purposes.

The system will provide security and configuration control on both base map data and controlled petroleum features of interest.

The system will provide for the interchange of data among the central facility and the various remote facilities.

The system will emphasize the use of commercial-off-the-shelf products in the implementation in order to reduce the costs associated with the hardware and software development efforts. Software development efforts should focus on the integration of the various COTS products functions through development of a point-and-click graphical user interface.

The system will conform to an open standard, scalable architecture in order to operate on various platforms and to incorporate future technologies.

Operational Scenarios

Base Map Data Entry

The PTS provides the analysts at the headquarters facility the capability to digitize base map data which resides on paper maps at various scales. The data may be digitized either using a digitizing table or scanning color separates of the maps and using raster/vector conversion software and on-screen editing tools.

Pipeline Data Entry

The EPA requires consortium members operating pipelines in the United States to submit pipeline data for pipes which meet certain specifications. At the remote facilities, analysts use PTS tools to input/update pipeline data using geocoordinates or base maps provided by the consortium. The analysts specify each feature's location, shape, and attributes. Entered data is then submitted by the remote facility's DBA to the consortium for inclusion into the centralized database.

Proximity Analysis

The EPA requires an environmental impact study to be performed on a proposed pipeline. The consortium has been tasked to identify all wildlife refuges and national parks which fall within five miles from the proposed pipeline. The analysts can generate a five mile buffer zone around the pipeline to assist in identifying effected refuges and parks.

Analyst-Defined Data Entry

An analyst has been asked to prepare a presentation of the results from a proximity analysis. The presentation should show local facilities owned by the company. PTS allows analysts to create new coverages which are stored in local workspaces. The analyst creates a building coverage and adds spatial features to the display which show the location of near-by facilities. This coverage, as well as the results from the proximity analysis (buffer zones), are imported into a map composition used for the presentation.

Data Query and Highlight

The EPA has identified the possibility of pipe failures (ruptures) for pipe sections manufactured by a company which went out of business in 1977. The ruptures have occurred for pipe sections installed above ground and under a pressure of more than 400 psi. The consortium has been tasked to locate any pipe sections meeting this criteria.

General Planning

A consortium member is proposing the implementation of a pipeline extending from Kuwait City, Kuwait to Cairo, Egypt. The company requires 1:250000 scale maps for the planning. These maps are provided by the consortium. The company enters a number of proposed pipeline routes as well as required service points along the pipelines.

Performance Parameters

(a)Drawing Speed

There are several properties of a vector layer which affect the time required to display the map features. These factors include the number of features, type of features, number of vertices, and the symbol representation. The performance factors for drawing and output are based on the test coverages: cull_pt (cross symbol), trns_ln (thin line), and veg_py (solid fill).

The system's response time for drawing the 'cull_pt' vector coverage in an 8.5 by 11 inch display window will not exceed 5 seconds.

The system's drawing rate for drawing the 'trns_ln' vector coverage in an 8.5 by 11 inch display window will not exceed 7 seconds.

The system's response time for drawing the 'veg_py' vector coverage in an 8.5 by 11 inch display window will not exceed 15 seconds.

The system's response time for drawing an 8-bit color raster in a 8.5 by 11 inch display window will not exceed 35 seconds.

(b)Output Rate

Map compositions which contain only vector data will be produced on color and black/white printers in less than 300 seconds.

Map compositions which contain any raster data will be produced on color and black/white printers at rates greater than 0.009 MB/sec.

(c)Retrieval rate

The system will query a coverage, returning vector features which satisfy a three operator boolean query, at a rate of 4 MB/sec.

(d)Login/initialization time

The response time from the initialization of the interface to selecting a coverage for drawing will not exceed 45 seconds.

(e)Optical Character Recognition (OCR) scanning time

The scan rate for the autofeed scanner. shall be 8 or more pages per minute.

(f)Scanning time

The large scale scanner shall scan an F–size map in less than 15 minutes.

Operational Deployment

The system consists of three types of operational sites: consortium headquarters (central), small facility (remote), large facility (remote).

Consortium Headquarters (Austin, Texas USA)

The headquarters facility is responsible for the input of base map data which is used by remote facilities for the input of pipe spatial data. Headquarters is responsible for maintaining and updating their central repository of pipeline and base map data. The analysts at headquarters perform data quality activities as well as configuration control on pipe data submitted by the remote facilities. General research projects and analysis tasks are also performed by consortium analysts at the headquarter facility. Figure 8 shows a list of activities, performed at the central facility, with the applicable performance factors and percentage of time the analyst utilizes the system for each activity. The percentage of time the user utilizes the system accounts for the percentage amount of time PTS is utilized to complete an assigned activity. All users require their own terminals.

Hours of Operation:	6am–6pm CST; if on–line activities exist with remote facilities, the system will be on–line during all hours of the remote systems operation with the exception of scheduled maintenance.	
Users:	(12)	Analysts/Researchers
	(1)	System Administrator
	(3)	Database Administrators
	(TBD)	Base Map Data Operators
Current Networks:	486 PCs on Novell Network.	

Current Software:	Word processing (MS Office), Email (Lotus Notes), GIS (PC Arc/Info)
Equipment:	~112 pieces (@ 7 pieces/user)

PTS Activities	*% Task Time PTS Utilized for Activity*	*Critical Performance Parameters*
Base Map Entry	100	drawing speed, login time
Text Entry	100	OCR scanning time
Pipe Data Entry	100	drawing speed, login time, scanning time
Analysis and Research	65	drawing speed, output rate, retrieval rate, login time
System/Database Admin.	80	drawing speed, retrieval rate, login time
Data Quality	90	drawing speed, retrieval rate, login time

Figure 8 Central Facility's PTS Activities

Large Remote Facility (Fairfax, Virginia)

This large remote facility has a need to input more than 5,000 pipelines into the system. These pipelines exist in a number of different regions in the world (six continents less Antarctica). The data entry group requires a large configuration of equipment. Figure 9 shows a list of activities, performed at the large remote facility, with the applicable performance factors and percentage of time the user utilizes the system for each activity. All users require their own terminals.

Hours of Operation:	6am–6pm EST	
Users:	(12)	Data Entry/Researchers (2 Divisions of six)
	(1)	System Administrator
	(1)	Database Administrators
Equipment:	~70 pieces (@ 5 pieces/analyst)	
Current Networks:	Sun (Sparc1, SparcIPC, Sparc2) on Ethernet	
Current Software:	Word processing (Framemaker), Email, GIS (Arc/Info)	

PTS Activities	*% Task Time PTS Utilized for Activity*	*Critical Performance Parameters*
Pipe Data Entry	95	drawing speed, login time
Analysis and Research	60	drawing speed, output rate, retrieval rate, login time
System/Database Admin.	50	drawing speed, retrieval rate, login time

Figure 9 Large Remote Facility's PTS Activities

Small Remote Facility (San Antonio, Texas)

This small remote facility has a need to input more than 400 pipelines into the system. These pipelines exist primarily in Europe and South America. The data entry group requires a small configuration of equipment. Figure 10 shows a list of activities, performed at the small remote facility, with the applicable performance factors and percentage of time the user utilizes the system for each activity. All users require their own terminals.

Hours of Operation:	6am–6pm CST	
Users:	(2)	Data Entry/Researchers
	(1)	System/Database Administrator
Current Networks:	386 PCs on Novell Network	
Current Software:	Word processing (MS Office) and Email (Lotus Notes)	
Equipment:	~15 pieces (@ 5 pieces/user)	

PTS Activities	*% Task Time PTS Utilized for Activity*	*Critical Performance Parameters*
Pipe Data Entry	95	drawing speed, login time
Analysis and Research	60	drawing speed, output rate, retrieval rate, login time
System/Database Admin.	20	drawing speed, retrieval rate, login time

Figure 10 Small Remote Facility's PTS Activities

Operational Life Cycle

The central system will be operational 01 May 1996. The remote systems will be operational on 01 June 1996. The system shall be designed for an expected life of five years.

System Usage

The system will be utilized over the hours specified in Figure 8, Figure 9, and Figure 10.

Effectiveness Requirements

Cost: The life cycle cost of the system shall be less than 28 million dollars over a seven year period.

Reliability: All servers shall have uninterruptable power supply support.

The mean time between failure (MTBF) shall be 20,000 hours.

The failure rate is 0.00005.

Maintainability: Preventive maintenance routines shall not exceed 1.5 hours.

The mean time to repair (MTTR) shall not exceed 6 hours.

The mean time between maintenance shall be one year.

Environmental Requirements

The system will operate in rooms with an average temperature of 72 degrees Fahrenheit. The temperature has a range from 65 to 85 degrees Fahrenheit. Any hardware requiring a special operating environment will need a room constructed to accommodate the hardware. The power available at the facilities in the USA is 110V/60Hz. The power at the Dover facility consists of 240V/70 Hz.

System Maintenance Concept

The system will consist of various COTS products integrated and accessed through a custom-built interface developed specifically for the consortium. The system's maintenance concept consists of two separate maintenance concepts: the Commercial Product Maintenance Concept and the Base Map/PTS Interface Software Maintenance Concept.

The commercial product maintenance concept provides two levels of support for hardware and software items provided by various commercial vendors. The maintenance tasks performed at each level are shown in Figure 11. The maintenance flow items include software bug reports and items needing repair. The supply flow items include software upgrades and patches, technical phone support, repaired or replaced items which needed repair.

The base map and interface maintenance concept consists of three levels of support. The maintenance tasks performed at each level are shown in Figure 12.

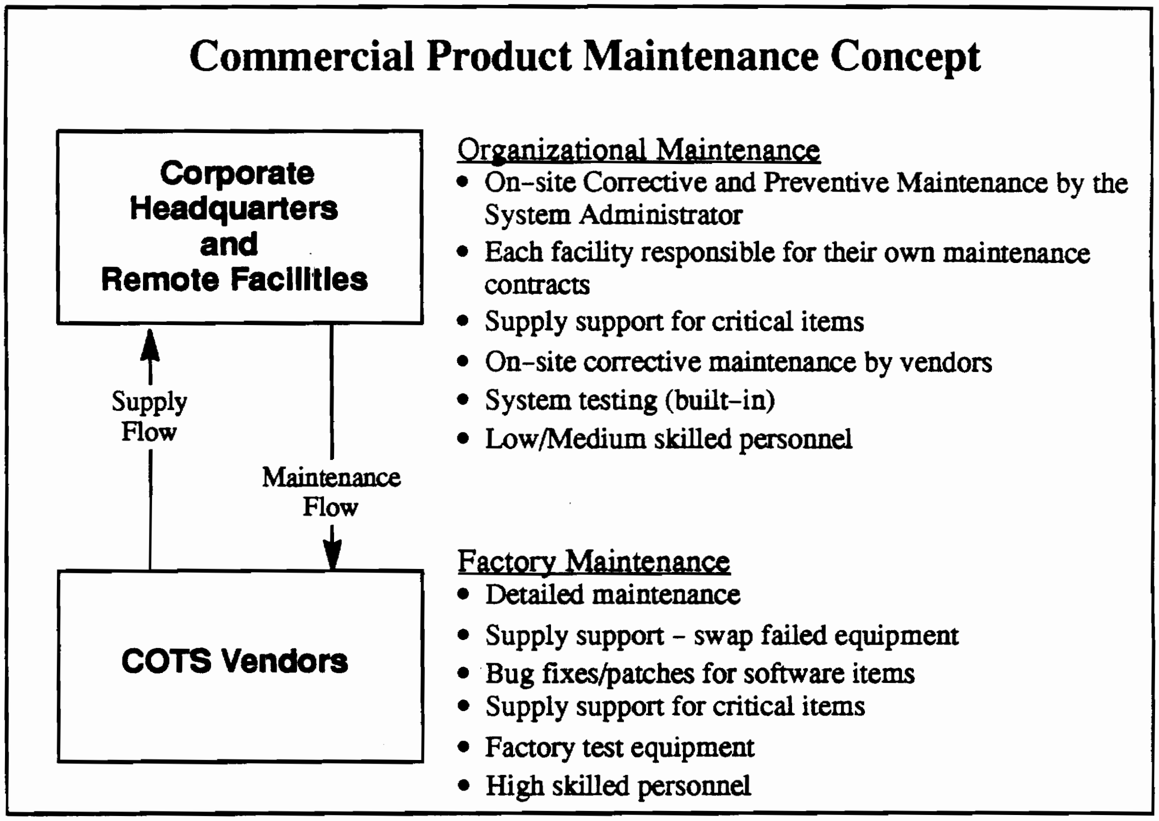


Figure 11 Commercial Product Maintenance Concept

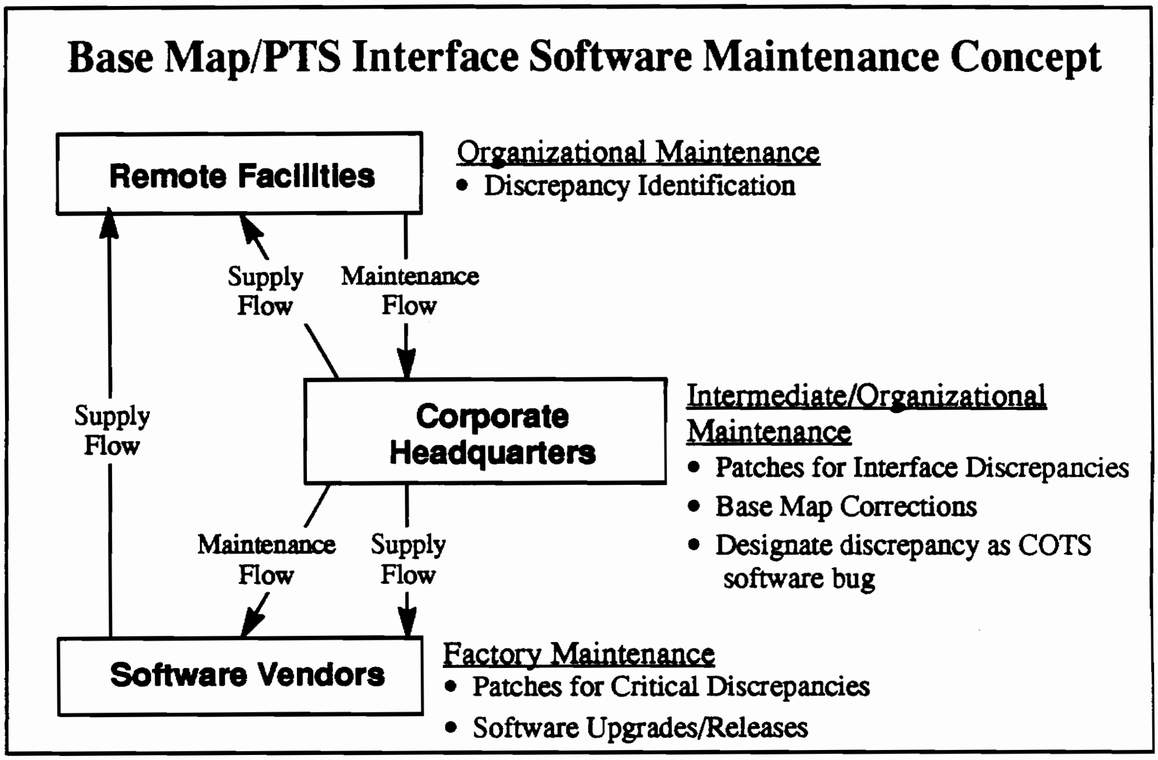


Figure 12 Base Map/PTS Interface Software Maintenance Concept

System Support

The planning of system support tasks and procedures throughout the system's life cycle is critical to maintaining the system's reliability requirements. System support activities for the PTS include maintenance planning (preventive), supply support of critical items, proper test and support equipment including tools, and training with documentation to assist system and database administrators in their support functions.

While the support planning begins with the maintenance concept, system administrators must work closely with the COTS vendors and their logistics staff throughout the life-cycle to ensure proper access to critical items and services necessary to restore system operations.

System administrators should keep preventive and corrective maintenance logs for all hardware and software items. In addition, any special handling procedures or tasks should be well documented.

System administrators have the following support tasks: configuration management of all hardware and software items, performance management of all network activities, preventive and maintenance on designated items (includes system backups), and security management of all user accesses.

Database Administrators have the following support tasks: configuration management of all spatial and attribute data under configuration control, loading and integrating new data sets and submitted pipeline data, database performance tuning, and database backups and restorations.

System Retirement

Due to a five year life span, an assessment should be made at the system's four year anniversary with respect to performing hardware and software upgrades to extend the system's life span. If the system will not undergo any extension in its life span, system administrators will need to perform the following tasks:

- Donate/transfer all hardware/software items which could serve other organizations.
- Recycle all recyclable material (paper, media, memory, etc.)
- Arrange for the disposal of items which require disposal.

Conceptual Configuration

Figure 13 presents the conceptual hardware configurations for the three PTS facilities. The central system consists of a central server which maintains all the data under configuration control (base maps and submitted pipeline data) and applications. The server will support a user group of 16 analysts, 25 conversion operators, and a support staff of 4 people. Each user requires his/her own terminal to access the server. The analysts have PCs on their desks which could be utilized to access the server using X server software.

The small remote facility requires a small server and 2-3 analyst terminals from which analysts may access the applications. The large facility has existing Sparc1, SparcIPCs and Sparc2 workstations for each analyst. These workstations may be utilized to access group and company servers for access to applications and data.

The remote systems may access the central server for data under configuration control and may submit pipeline data across communication links to the central facility.

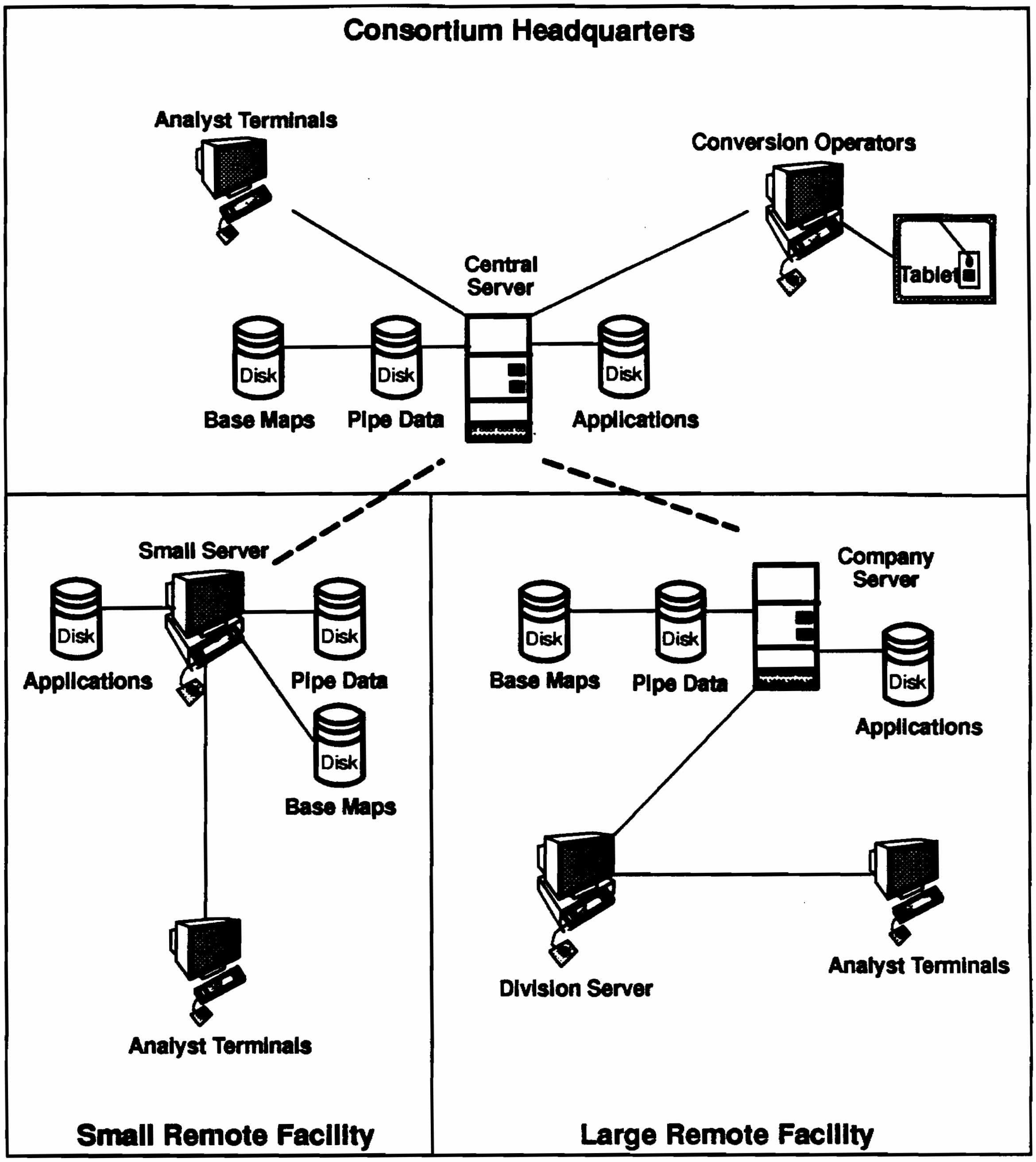


Figure 13 Conceptual Configuration for the PTS

Operational Requirements

The following sections present the operational requirements with respect to the COTS products (system), the base maps, and the customized interface used to access the various COTS products as well as the GIS's functionality. The operational requirements are organized based on these three aspects of the system:

- System Requirements
- Base Map Library Requirements
- System Interface Requirements

System Requirements

<u>VECTOR PRODUCTS</u> The system shall store and display world vector maps with scales of 1:3M (ESRI's ArcWorld) and 1:1M (Digital Chart of the World). These products are used to assist the users in locating their areas of interest (AOI).

<u>LARGE SCALE AREAS OF INTEREST</u> The system shall store spatial information for large-scale maps (i.e. 1:50,000 or less) independent of the a base map library. Polygon features shall be created within base maps to represent the map extent of large-scale areas of interest.

<u>TOPOLOGICAL STRUCTURING</u> The system shall be able to convert digitized strings (i.e., points, polygons, arcs) into defined topological structure.

<u>VECTORIZED INPUT</u> The system shall provide for a capability to input previously digitized data contained on various digital storage media. The following format shall be supported:

- Arc/Info

<u>COMPRESSION</u> All raster and vector data utilized by the GIS will be stored uncompressed.

<u>VECTOR FORMAT</u> The system will store vector data in Arc/Info format.

<u>RASTER FORMAT</u> The system shall store raster images in a tiff format supported by the GIS software's image viewer.

<u>MAGNETIC MEDIA</u> The system shall provide the operator with the capability to read magnetic media into the data base. The system shall support the following magnetic media formats:

A. 3.5" floppy disks

B. 8mm tapes

C. 1/4 tapes

System Requirements

OPTICAL DISK The system shall provide the operator with a capability to read information stored on optical disks in CDROM format.

REWRITABLE OPTICAL DISK The system shall provide the operator with a capability to read and write coverage, attribute, and map composition data to a 1.3 GB rewriteable optical disk.

OCR SCANNING The system shall provide a capability to digitize textual documents, via an OCR scanner, for entry into a system data base or file.

SCANNING The system shall provide a capability to scan and store binary, grayscale, 24-bit color hardcopies (paper maps).

MAP SCANNING SIZE The system shall support the scanning for hardcopies sized for 8.5" by 11" to 44" by 32" (F-scale maps) for large map scanners.

MAP SCANNING RESOLUTION The system shall provide a capability to scan and store images at resolutions ranging from 120 dots per inch (dpi) to 500 dpi for the large map scanners.

DESKTOP SCANNING SIZE The system shall support the scanning for hardcopies sized for 8.5" by 14" for the desktop scanners.

DESKTOP SCANNING RESOLUTION The system shall provide a capability to scan and store images at resolutions ranging from 120 dots per inch (dpi) to 400 dpi for the desktop scanners.

COLOR SEPARATES The system shall provide the capability to produce individual red, green, and blue scans using the large scale scanner. The system shall store the scanned separates in a format compatible with the raster/vector conversion software.

DIGITIZATION The system shall provide for free format digitization of cartographic information or printed map sheets. The methods for digitization shall include:

A. Entering coordinates by keyboard

B. Interactive Raster to Vector registration

A. Digitizer Tablets

B. Digitizer Table (Lighted)

C. Terminal Mouse

D. Raster/Vector Conversion Software

EDITING TOOLS The system's interface shall provide the capability to crop the image's size and modify/erase pixels within a stored raster image.

HORIZONTAL INTEGRATION The system shall be capable of horizontally integrating (edge matching) map input data with existing data base contents. Additionally, the system shall provide a capability for horizontal integration of individual map sheets with the existing data base contents to create a new map.

System Requirements

VERTICAL INTEGRATION The system shall be capable of vertically integrating additional layers of map input data with existing data base contents. Additionally, the system shall provide a capability for vertical integration of data from more than one layer at a time to create a new map.

SECURITY STANDARDS The system shall be structured to provide security according to individuals, groups, and program privileges.

STORAGE CAPABILITY The system shall be capable of storing a database whose geographic extent incorporates all areas of interest.

FLEXIBLE ORGANIZATION The system shall permit the restructuring of the database so that attribute fields may be added, redefined, or deleted.

DATA BASE RELATIONSHIPS The system shall be capable of relating cartographic data with textual data.

DATA BASE PARTITIONING The system shall provide the capability of dividing the data base into discrete segments representing different geographic areas of interest.

DATA BASE LOCATION The system shall provide for the storage of portions of the data base in a central location (e.g., file server), or a specified location (e.g., an analyst's workstation).

DATA ACCESS Data base contents may be accessed by all operators, subject to established security procedures.

TEXT SEARCH The system shall provide the operator with the capability to scan multiple text files by using a free format search process. The response time shall not be greater than that specified by the manufacturer of the selected COTS product.

OUTPUT DEVICE The system shall provide a high quality device such as an electrostatic printer to provide graphics and textual output in different dimensions (8"x11" to 30"x40").

NETWORK INTERFACE Each PTS workstation shall have the capability to be networked to other PTS workstations in a facility.

PROCESSOR The workstations and servers shall be based upon a 32 bit processor.

COLOR DISPLAY The system shall provide color graphics display for the presentation of cartographic information. The display device shall be capable of displaying a minimum of 16 different colors.

DISPLAY SIZE The monitor for the color graphics display shall be, at a minimum, 16" measured diagonally.

DISPLAY RESOLUTION The color monitor shall be capable of displaying a minimum of 1152 (horizontal) by 900 (vertical) pixels.

MOUSE The system shall provide for the input of operator commands through the use of a mouse to select options from displayed menus.

System Requirements

KEYBOARD The system shall provide a keyboard for operator impact of parameter values, and as an alternate to the mouse for input of commands.

DIGITIZING TABLET/TABLE Each facility shall be provided with a digitizing tablet/table.

ON-LINE STORAGE The system shall provide on-line storage for the system's spatial and attribute databases.

Base Map Library Requirements

GENERAL USE The base maps shall be used for locating and positioning a new pipe feature into the spatial database. The base maps shall be used within a composition mode to provide proximity presentations of pipe features to map features represented in the base maps.

SPATIAL DATA The base maps shall provide the locations of features within the following cartographic themes:

- transportation features: paved roads and highways, airports, and railroads.
- land/ocean boundaries
- political boundaries: country, province, districts, and counties borders
- drainage features: rivers and inland bodies of water.
- wildlife refuges, parks, and environmental sensitive zones.
- city features: cities with populations greater than 1000 people.
- petroleum features: industrial areas, wells, storage facilities, and buildings.

BASE MAP SCALES The base map data shall consist of maps at scales of 1:1000000, 1:250000, and 50:000.

PROJECTION The base map data for areas of interest shall be stored in the georeference coordinate system (decimal degrees).

AREA OF INTEREST The system shall provide base map data for the areas of the earth between the latitudes of 80N and 60S.

ATTRIBUTE DATA The base maps shall identify a map feature's theme, type, and if available, the official name.

COMPRESSION All base map data utilized by the GIS will be stored uncompressed.

VECTOR FORMAT The system will store vector base map data in Arc/Info format.

RASTER FORMAT The system shall store raster base map data in a tiff format supported by the GIS software's image viewer.

System Interface Requirements *(Display)*

SELECTION OF DATA SOURCES The system's interface will allow the analyst to select and view any data source which falls within an analyst's map extent of interest. The data sources include vector libraries and registered images.

MAP DATABASES The system's interface shall provide each operator with on-line access to a large volume of base maps of varying scales and projections for different areas of interest of the world.

PAN/ZOOM The system's interface shall provide the capability to modify the extent of the display window using Fullview, Pan, Zoom-In, Zoom-Out, Last Extent, Box, and Redisplay.

MAP DISPLAY The system's interface shall be capable of displaying operator selected cartographic data on a color graphics terminal.

DISPLAY FORMATTING The system's interface shall permit the operator to format displays of cartographic data.

ADD COVERAGES The system's interface shall provide the operator the capability to add vector and registered images, which reside in other libraries or local workspaces, to the display.

REORDERED DRAW ORDER The system's interface shall provide for the capability to reorder the display of coverages within a the coverage type, i.e. point, line, polygon.

SYMBOLIZATION The system's interface shall permit the operator to specify the method for symbolizing features with different attributes (e.g., color, pattern, font).

DISPLAY SYMBOL ASSIGNMENT For the coverage types specified by the consortium, the system's interface shall display default symbols for coverages of the same type (e.g., roads, buildings, vegetation, etc.)

SYMBOL EDITOR The system's interface shall provide a capability to edit existing symbols or to create new symbols which may be added to the symbolset in the local workspace. New symbols may be added to the master symbolset through established configuration controls.

DISPLAY of ATTRIBUTE DATA The system's interface shall be capable of displaying all attribute information related to the cartographic features selected by the operator.

DISPLAY of RELATED TEXTUAL DATA The system's interface shall be capable of displaying textual information associated with the cartographic features selected by the operator.

VIEWS The system's interface shall create and load views which allow the operator to save the set workspaces and coverage selections, the symbol settings, query results, and the display's map extent.

System Interface Requirements *(Analysis)*

ATTRIBUTE BASED RETRIEVAL Retrieval of features based upon attribute values shall be permitted. These retrievals shall be based either upon general feature description (e.g., railroads), or a specific feature class (e.g., type of roads).

SELECTION SET DEFINITION The system's interface shall permit the refinement of a total selection set into defined subset(s).

RELATIONAL OPERATORS The following relational operators shall be provided to assist in the query of cartographic feature attributes:

- = Equal
- ^= Not Equal
- > Greater Than
- < Less Than

BOOLEAN OPERATORS The system's interface shall provide the following Boolean Expressions to assist in the query of cartographic feature attributes:

- And
- Not
- Or

DATA EXTENSION The system's interface shall be capable of extracting a subset of features out of the cartographic data base to create a smaller working data set.

MEASUREMENT TOOLS The system's interface shall provide the following measurement tools:

- Length Calculation - computing the length of a linear feature
- Coordinate Calculation - determining longitude/latitude coordinates of a selected position
- Closed Curve Calculation - determine the area included by a closed curve

System Interface Requirements *(Spatial Editing)*

REGISTRATION The system's interface shall provide a means to register a map sheet to database coordinates for digitization. The method for registration shall include:

CREATE COVERAGE (THEME) The system's interface shall provide a means define a new coverage.

DIGITIZATION The system's interface shall provide for free format digitization of cartographic information or printed map sheets. The methods for digitization shall include:

A. Entering coordinates by keyboard

B. Interactive Raster to Vector registration

A. Digitizer Tablets

B. Digitizer Table (Lighted)

C. Terminal Mouse

D. Raster/Vector Conversion Software

CREATE FEATURE (THEME) The system's interface shall provide a means to create arc, point, region, and annotation features within a coverage. The feature can be entered using coordinates given at the keyboard (or coordinate file) or by positioning the mouse's pointer on the display.

ERROR DETECTION The system's interface shall provide a means to detect digitizing errors (e.g., dangling nodes).

ERROR DISPLAY The system's interface shall be able to display detected errors to an operator to facilitate corrections.

LINEAR FEATURE GENERALIZATION The system's interface shall provide the operator with a capability to generalize lines to reduce the amount of detail.

LOCATIONAL EDITING The system's interface shall provide locational editing capabilities for editing topologically structured data (e.g., move, copy, add, delete).

ANNOTATION PLACEMENT The system's interface shall permit the operator to specify the placement and orientation of feature annotation.

SPATIAL DATA PRESERVATION The system's interface shall provide safeguards to avoid unauthorized manipulation or accidental destruction of spatial data under configuration control.

System Interface Requirements *(Attribution)*

ATTRIBUTION The system's interface shall provide a capability for the user to interactively assign attributes, which are stored in relation databases, to topologically structured digitized data residing in the local workspace.

ATTRIBUTE EDITING The system's interface shall provide for operator editing of attribute data that resides in the operator's local workspace.

KEYSTROKING The system's interface shall provide the operator with text input capabilities via keystroking.

VALUE LISTS Where applicable, the system's interface shall provide the operator with a list of available field values that can be selected to automatically populate the data field.

VALIDITY CHECKING The system's interface shall provide a capability to automatically check user supplied input for conformance to valid ranges of attribute values.

ATTRIBUTE DATA PRESERVATION The system shall provide safeguards to avoid unauthorized manipulation or accidental destruction of attribute data under configuration control.

TEMPLATES The system's interface shall provide the capability to create attribute fields for coverages based on a predefined attribute template.

DATABASE ACCESS The system's interface shall provide all users, subject to established security procedures, the capability to access the database attributes.

DATABASE QUERIES The system's interface shall provide an tool external from the GIS which can perform defined and ad-hoc queries on the attribute databases.

REPORT GENERATION The system's interface shall provide the capability to produce formatted reports of retrieved data from the attribute databases.

System Interface Requirements *(Utilities)*

REWRITABLE OPTICAL DISK The system's interface shall provide the operator with a capability to read and write coverage, attribute, and map composition data to a 1.3 GB rewriteable optical disk.

FEATURE INTEGRATION The system's interface shall provide the capability to associate textual information with cartographic features.

PROJECTION CONVERSION The system's interface will provide the user with the capability to convert their own map data from any of the following projections into any other one:

- TRANSVERSE - Transverse Mercator
- UTM - Universal Transverse Mercator
- GEOGRAPHIC - Longitude and Latitude
- LAMBERT - Lambert Conformal and Conic

TEXT DATA BASES The system's interface shall provide the operator with on–line access to a large quantity of formatted and free text data bases drawn from a variety of sources.

TEXT OUTPUT The system's interface shall provide the operator with the capability to display textual data.

IMAGE RETRIEVAL The system's interface shall provide the operator with a capability to retrieve frames of imagery of a specific area of interest from their working imagery file, and have it displayed at the operator's workstation.

MAP SHEETS The system's interface shall be capable of outputting data in the form of high quality colored cartographic map sheets for reports and presentations.

FORMATTING The system's interface shall permit the operator to format map outputs with labels, legends, and notations.

LEGEND GENERATOR The system's interface shall provide a capability to generate a legend in a map composition for any preexisting coverage loaded into a map composition.

GRAPHICS OUTPUT The system's interface shall provide the operator with the capability to produce color graphics in the form of hardcopy or transparencies.

Project Schedule

The project's schedules for both the system and the base maps are shown in Figure 14 and Figure 15. The system's schedule consists of four major project reviews: the conceptual design review, the preliminary design review, the detail design review, and the operational readiness review. The base map's schedule consist of two major reviews: the preliminary design and detail design reviews.

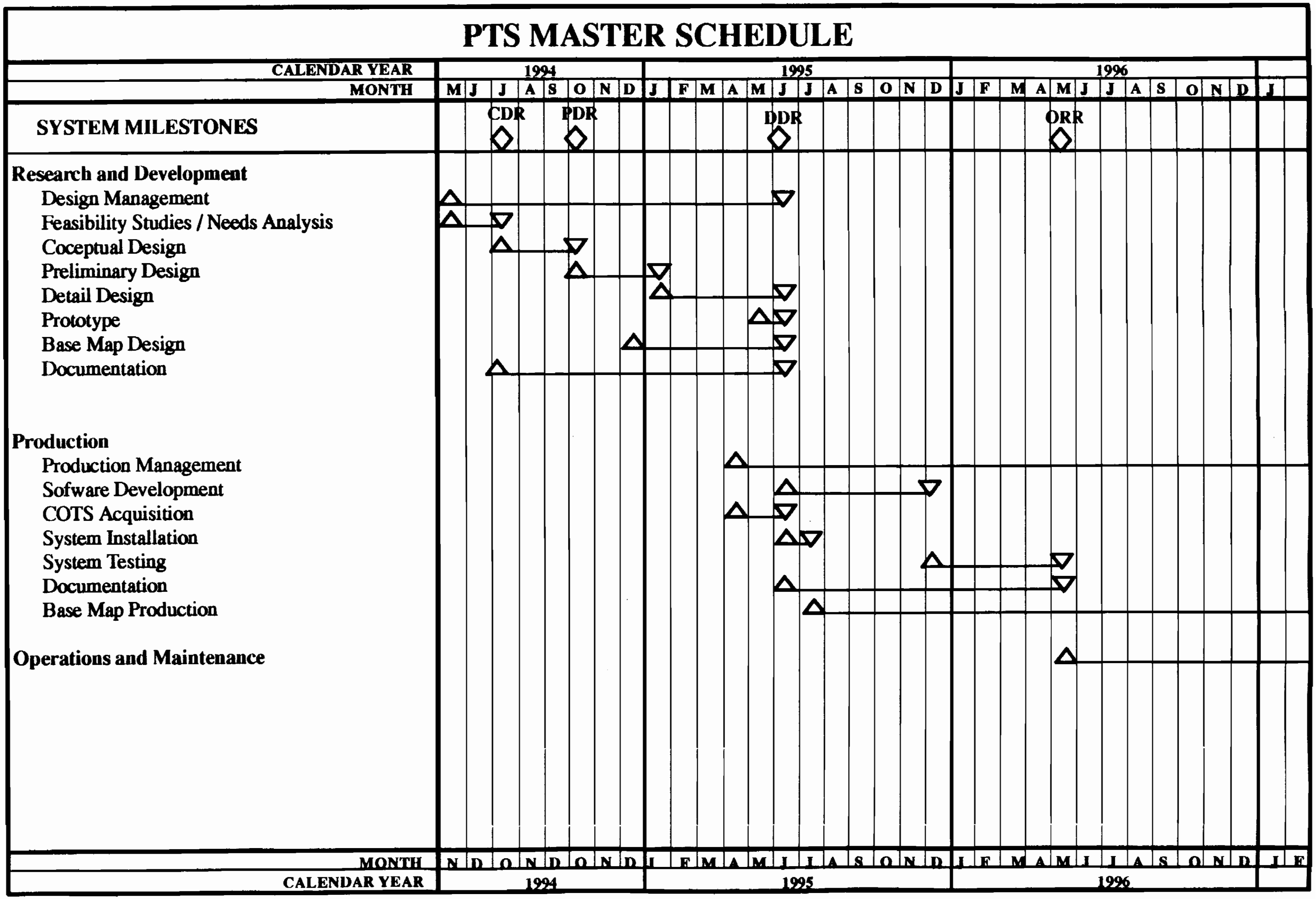


Figure 14 PTS Master Schedule

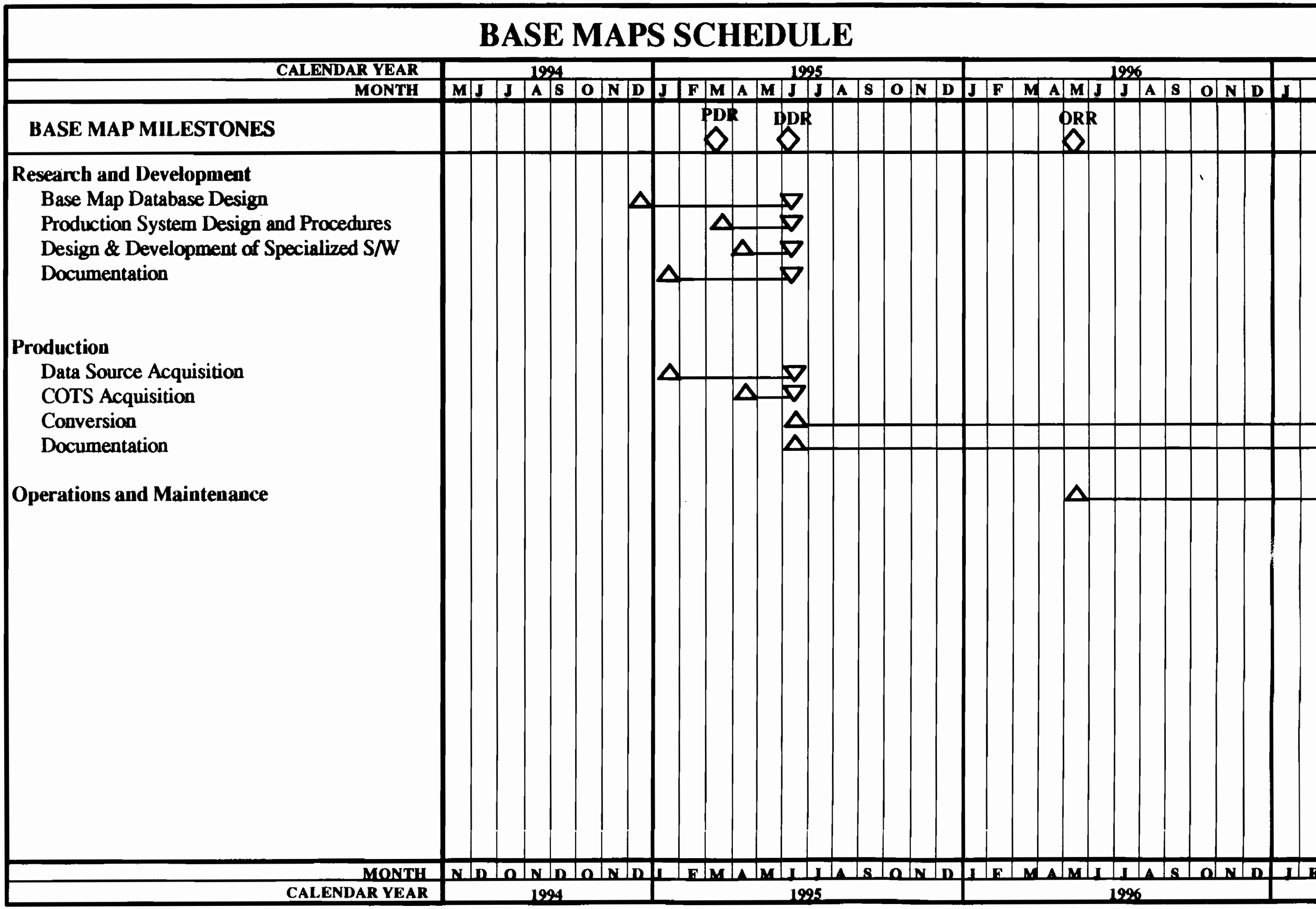


Figure 15 Base Maps Schedule

SECTION 3

PRELIMINARY DESIGN

The preliminary design takes the system specifications, assumed to be approved by the conceptual design review board, and begins to synthsize and define the system using qualitative and quantitative parameters. The activities from which the preliminary design evolves from include functional analysis, functional allocation of operational requirements, and trade-off and optimization.

Functional Diagraming of the System

Functional diagraming is used to define the PTS into its functional components. The intent of diagraming the system's functional components is to facilitate the design by identifying the numerous functional interfaces and by presenting the system in a logical structure. The diagrams begin with the top level, Figure 16, which describes the gross operations of the system. From the top level, each functional block is broken down in more functional activities which in turn are broken down further into more functional activities. Using the operational and maintenance concepts, the functional diagrams continue to be broken down into subfunctions until the subfunctions can be identified with particular needs (hardware, software, people, data, etc.) of the system. Figure 16 through Figure 25 show the first three functional levels of the operations activities. Figure 26 shows the first level maintenance functions for the evaluation of system operations.

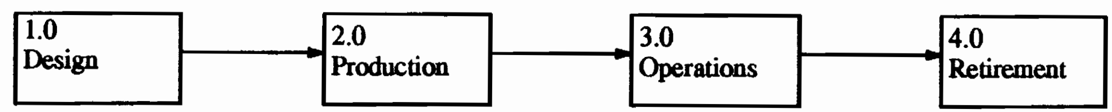


Figure 16   Top Level Functional Diagram for the Pipeline Tracking System

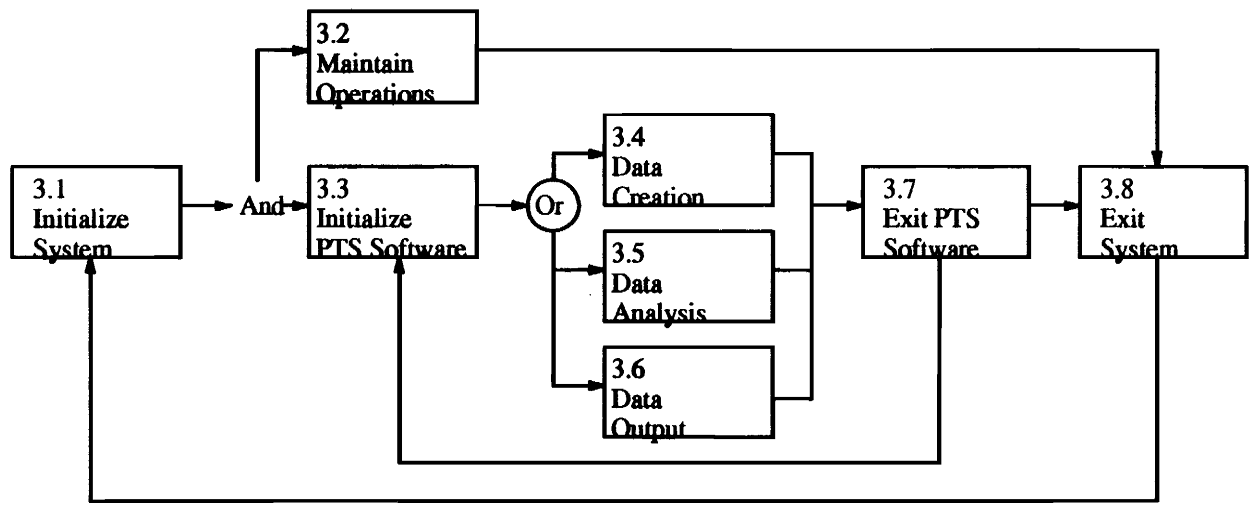


Figure 17 Second Level Functional Diagram for Operations

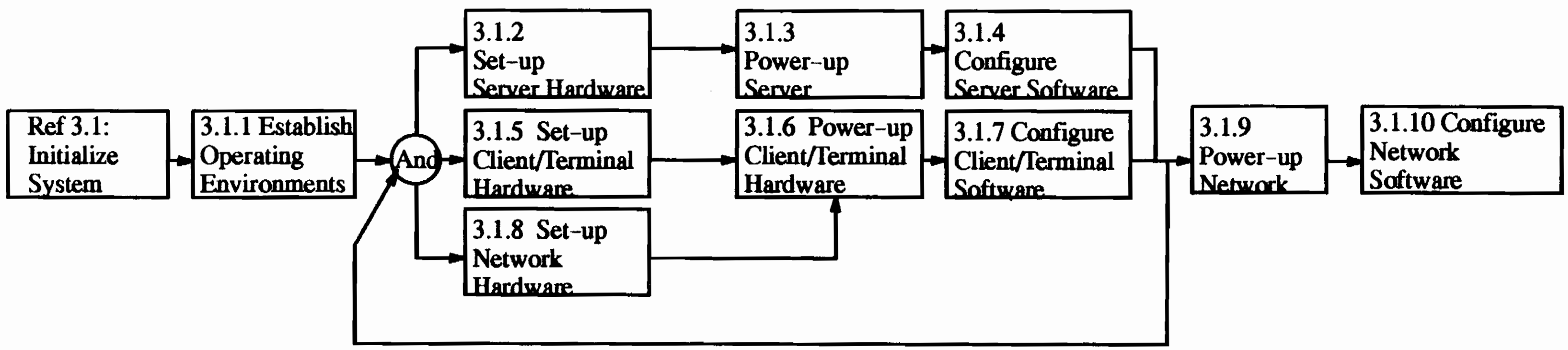


Figure 18 Third Level Functional Diagram for Operations (Initialize System)

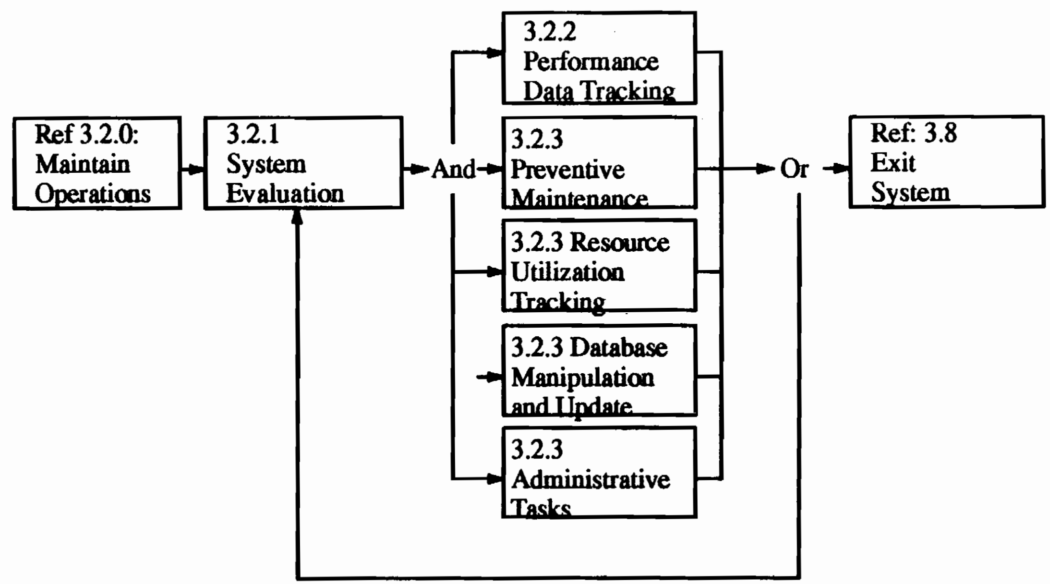


Figure 19 Third Level Functional Diagram for Operations (Maintain Operations)

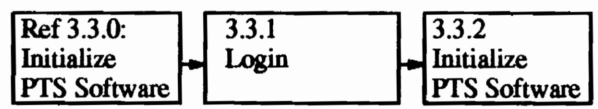


Figure 20 Third Level Functional Diagram for Operations (Initialize PTS Software)

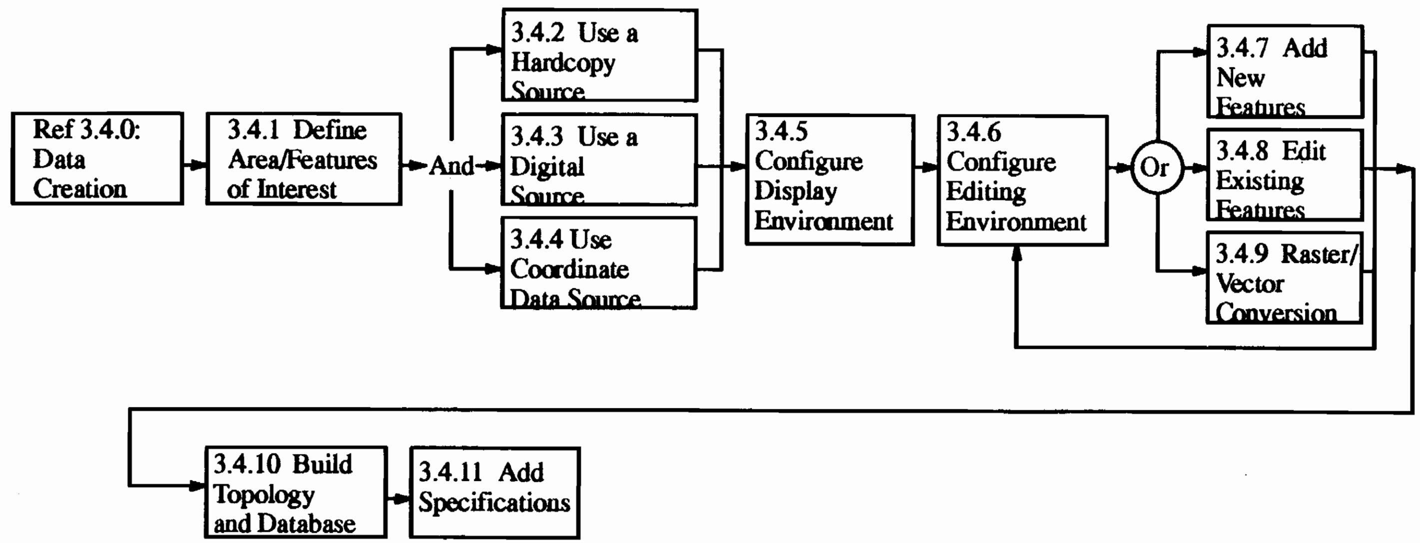


Figure 21 Third Level Functional Diagram for Operations (Data Creation)

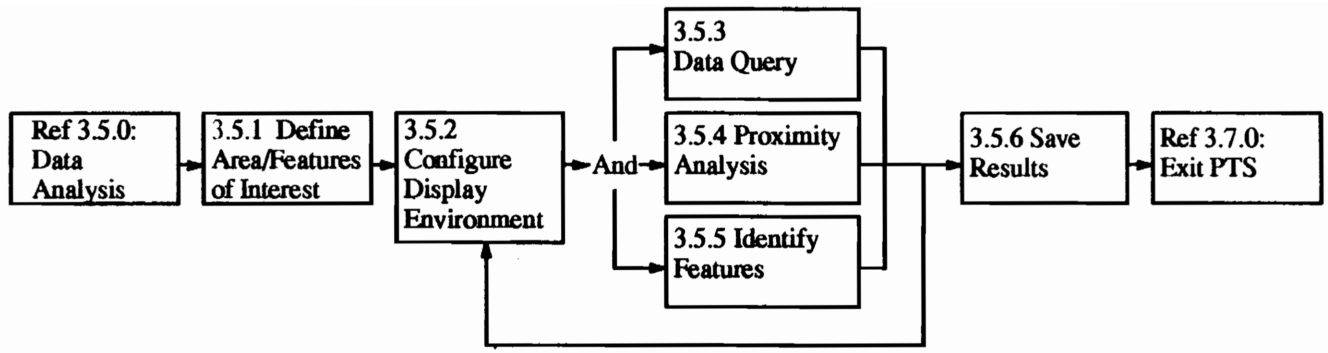


Figure 22 Third Level Functional Diagram for Operations (Data Analysis)

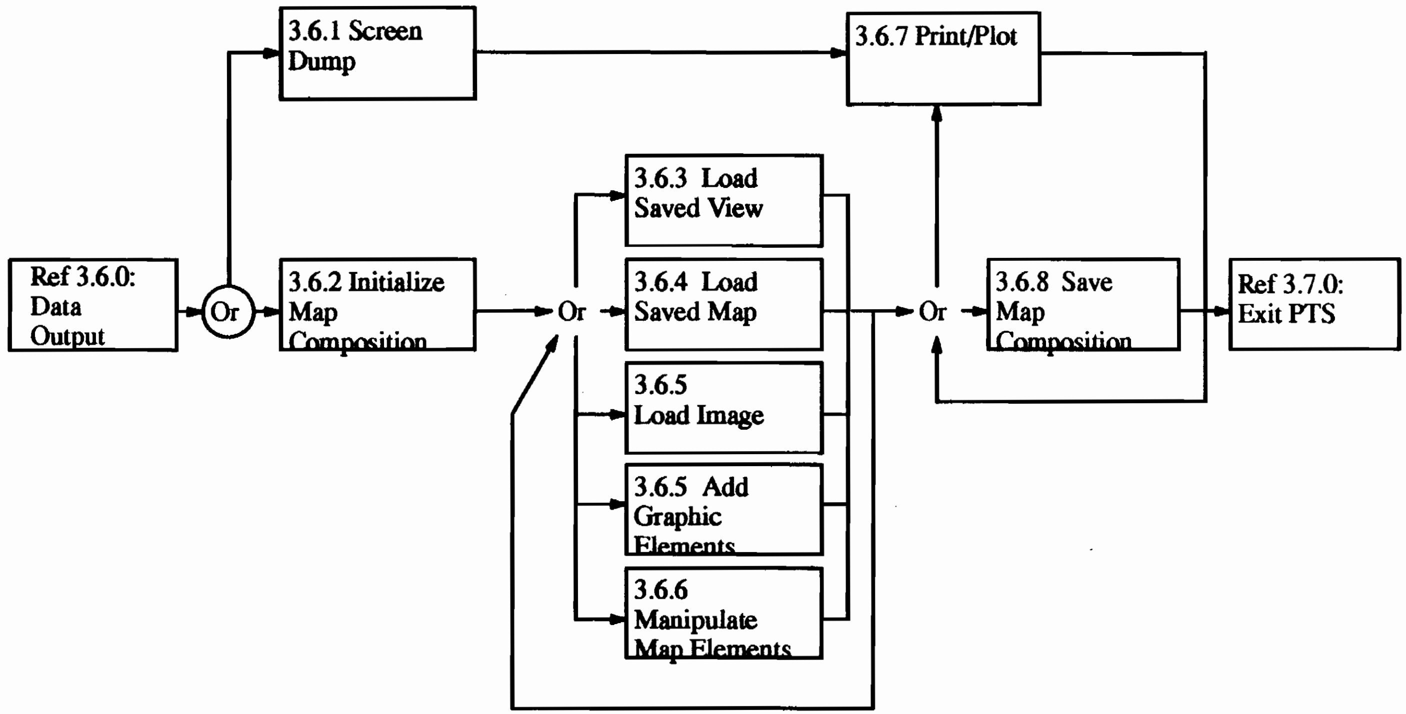


Figure 23 Third Level Functional Diagram for Operations (Data Output)

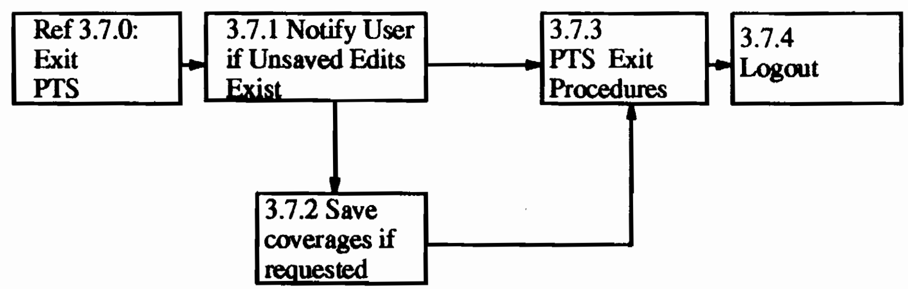


Figure 24 Third Level Functional Diagram for Operations (Exit PTS)

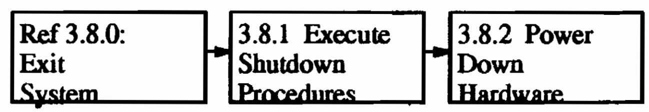


Figure 25 Third Level Functional Diagram for Operations (Exit System)

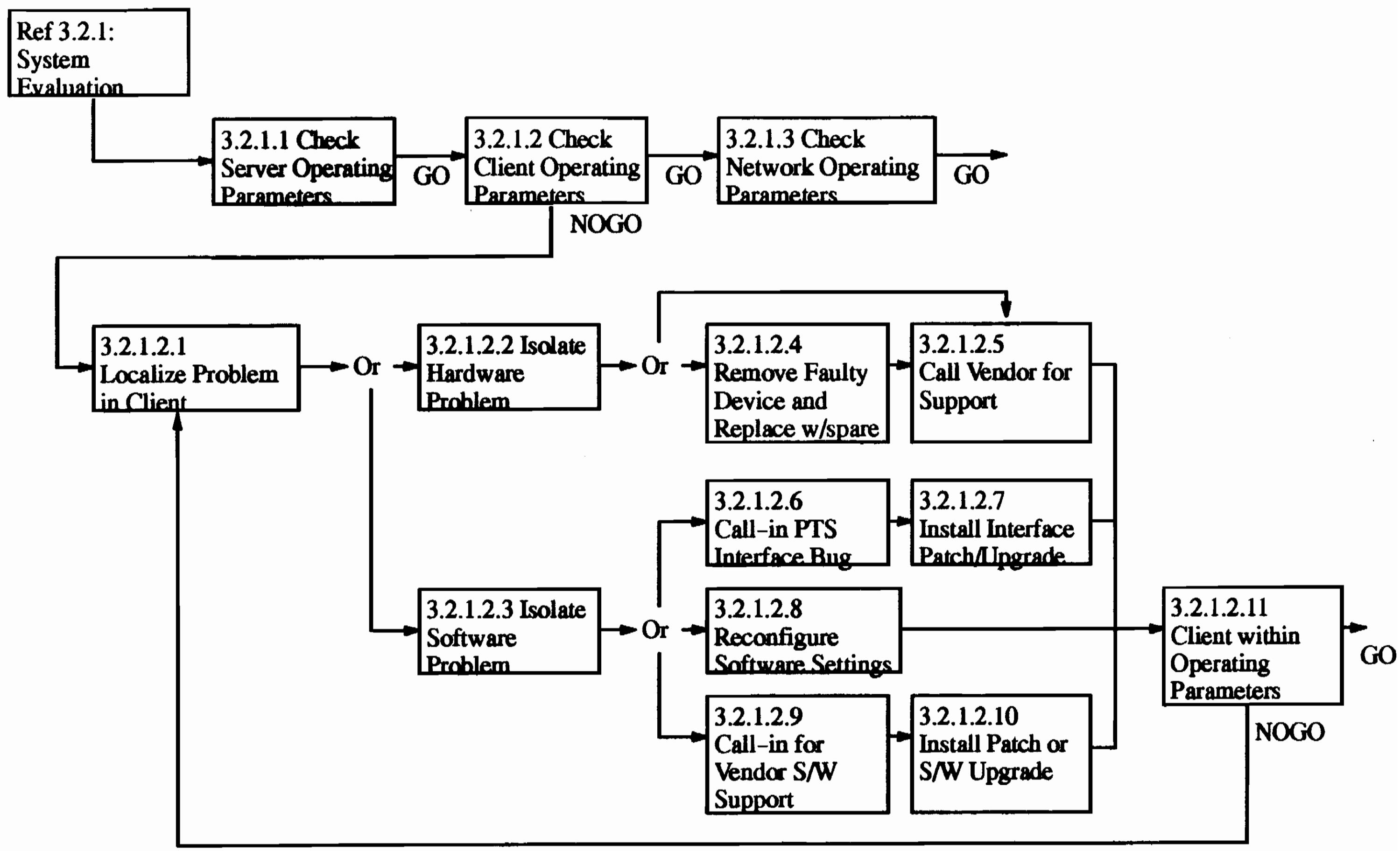


Figure 26 First Level Maintenance Flow Diagram for System Evaluations

Functional Allocation of Requirements

The allocation of requirements is used to further define the functional components of the PTS in terms of their specific performance requirements. The intent of allocating requirements is to facilitate the design by identifying performance and maintance properties of the numerous functional components. Figure 27 show the allocation of requirements for the PTS.

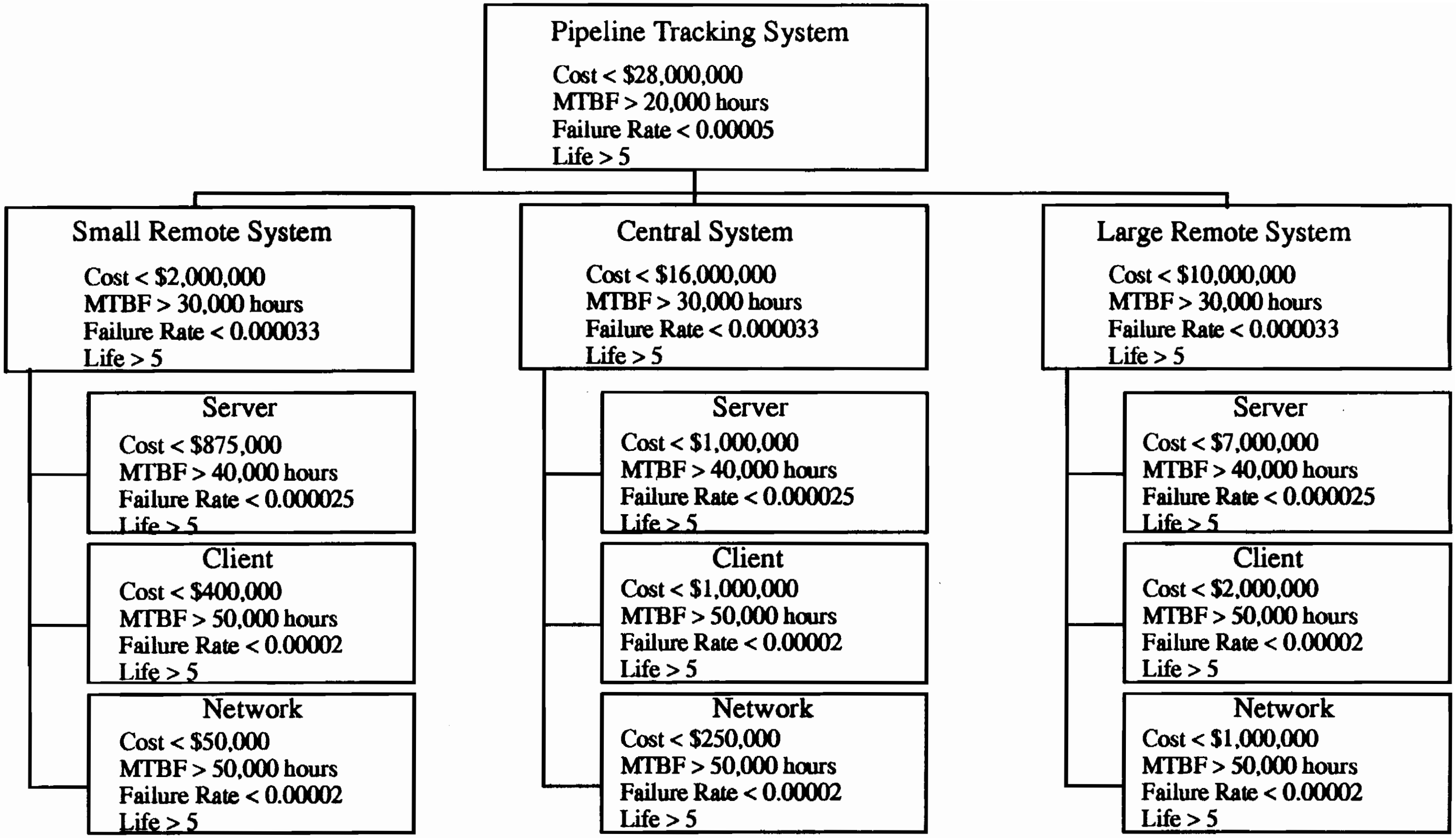


Figure 27 Functional Allocation of Requirements

Evaluation of Raster and Vector Base Map Implementations

The most common methods for entering new spatial data into the system include using a digitizer cursor on a digitizing table, a keyboard to enter latitude/longitude coordinates, or a mouse to digitize features on the screen. While all three methods are utilized within the system, the majority of the pipelines are to be entered into the system using the mouse to place features directly on the screen. In order to digitize features directly on the screen, an analyst requires base map data as a reference to locate where features are to be placed. For example, an analyst must digitize a pipeline, 150 feet from Interstate 45, that runs parallel to the interstate from mile posts 240 to 250. In this case, an analyst would require a base map which shows Interstate 45 with sufficient resolution to distinguish a 150 foot separation and individual mile posts.

Most of the pipe data is entered by analysts at various remote facilities (consortium members: petroleum companies). Once entered, this data is first consolidated within the company. Then the company's DBA submits the pipelines to the central facility where the pipelines are consolidated into a single spatial database. If the remote facilities enter their data using different base maps, consolidation of the data become more difficult. These difficulties include locational variances among various map products, map inaccuracies, different projections, and the unknown quality of the base map source. By providing the remote facilities with a standard set of base maps created and controlled by the consortium, the consolidation of pipe data into a single spatial database will be possible. Since most of the pipe data is located using these base maps, these maps have a critical role in the creation of a centralized repository of pipeline data. The accuracy of added pipeline features can only be as good as the accuracy of the base map upon which the features are added.

The consortium has access to a large library of hardcopy maps for the entire region of interest (land area extending from latitude 80N to 60S. The entire map set exists in three scales: 1:1,000,000; 1:250,000; and 1:50,000. These maps are the data sources from which base maps database for PTS are created.

The base maps can be maintained in either raster or vector format. The specific concerns with regard to the life–cycle costs are the data design, production, operations and maintenance, and disposal costs. In addition, both formats have their own effects on the system's network and workstation resources.

The following section provides an analysis of resources required to operate PTS with either a raster or vector base map database. This section evaluates the following items:

- Usage/Presentation
- Data Storage
- Display Performance
- Network Performance
- Life Cycle Costs

Usage/Presentation

Vector Base Maps

Vector data consists of coordinate pairs located in a Cartesian coordinate system. For the PTS system the base map data will be displayed using the georeference (longitude/latitude) system for the coordinate system. Features on the Earth's surface are mapped onto a flat, two-dimensional map as points, lines, and regions (polygons). Points are represented as single coordinate pairs. Lines are represented as strings of coordinate pairs. Polygons are represented as strings of coordinate pairs which define a closed area on a map. (3)

Vector data provides a feature-based representation of map features and is more accurate in representing a feature's actual physical shape than using raster pixels. In addition, vector features can be given unique ids which may be used to associate features to individual database records.

Vector data is organized into layers or themes. For example, a transportation layer contains all the roads and railroads within a defined map extent. This layer structure allows analysts to display only the data that is of interest.

To create a vector base map database, the hardcopy maps are inputted into PTS by using a large-scale scanner to produce three grayscale separates which are stored in raster. These scans are made with a high degree of resolution (500 dots per inch) in order to assist the raster-to-vector conversion software's identification of linear borders. To create a vector layer of features, the conversion operator uses raster-to-vector conversion software to produce a first draft of the vector layer. The output from the software is then cleaned up (removing non-theme features, digitizing features not recognized by the software, and populating gaps and holes in the layer based on the scanned separate). Following the clean-up, attributes are transferred and assigned to individual vector features using the hardcopy map and other sources as references.

By making presentations of pipe data using vector base maps, an analyst can control the display of the features of interest. This makes for charts that are focused and direct in the presentation of data. However, the analyst must be careful to provide sufficient locational information; otherwise, the audience may have trouble locating where the features exist on the ground.

Raster Base Maps

Raster data consists of regularly spaced cells called pixels. These pixels are organized in rows and columns and represent rectangular locations. Each pixel is assigned a reflectance value. The value can be interpreted and displayed on a monitor as a color or gray-scale through the use of a colormap.

Raster data is used for geographic feature representation and image picture databases. For geographic feature representation, the raster format provides a good representation of features which have gradual boundary transitions such as soil type and elevation. (4) The raster format will not be used to represent the consortium's pipeline features since these features have distinct edges and can be better represented as vector features. However, implementing the base map database in a raster format results in a large image database which provides an effective reference for the entry of pipeline data.

A raster base map is organized as a single image. By displaying a raster base map, the analyst is provided with all of map's information. Since the base map assigns pixel values for all the pixels in the display window, the display of multiple raster maps for the same map extent results in the view of the last image displayed. The vector format supports overlays of base maps from different map sets within the same map extent.

To create a raster base map database, the hardcopy maps are inputted into PTS by using a large-scale scanner to store the image in a raster format. The stored images are then registered to the base map database's coordinate system. Following the registration, scanned images are rectified (scaled and rotated) to best fit the coordinate system (ie. latitude/longitude) of the display window. If the paper map uses a coordinate system other than geographic, which is usually the case, the rectification process performs its best fit to the registration points.

Presentation and reference problems occur when neighboring images are displayed within the extent of the display window. When displaying an extent larger than the raster image, neighboring images must be displayed. This can result in long drawing times, overlapping images, and horizontal mismatches to neighboring images especially if the original paper map is rectified to fit another coordinate system. An example is viewing a map with a Universal Transverse Mercator projection in a georeference coordinate system. The overlapping images can result in pipeline digitizing errors; the rectified image adjusts well to selected registration points but may be off in other points due to the distortions that are created when a paper map is rectified. Maintenance costs occur when database administrators as well as analysts need to spend time rerectifying base maps images to obtain better georeferences for specific areas of the base maps.

Presentations of pipe data using a raster base map provide the audience with a wealth of information with regard to where the pipelines resides. However, with the inclusion of the raster base map in the presentation, the results from the analysis or query can become 'lost' as the presentation becomes 'busy'. Presentation problems can also occur when the map extent extends across base maps whose hardcopy maps are different products having different feature representations and background colors.

Presentation Summary

The vector format provides the consortium with the greatest level of flexibility for the purpose of feature tracking. The vector format provides flexibility in the display of selected features rather than all features which allows the analyst to spend more time analyzing rather than waiting for the entire base map to draw. If an analyst needs to show selected features from a raster base map in a map composition, those raster features need to be digitized by the analyst. In addition, the analyst working with vector data may change the display from one coordinate system to another. Since the hardcopy base maps will come from different sources, the raster base map database's map feature representation may change with the extent while the vector format would provide for a more standardized representation of data.

Maintenance costs would be reduced using the vector format since updates could be performed within the individual layers. For updates to a raster base map, an entirely new map sheet would have to be created, procured, and scanned. In addition, updates to raster base maps could not be made until new editions to the hardcopy maps are produced while changes to the vector can be performed with a smaller turnaround time. The rectification process of scaling and rotating raster base maps from other coordinate systems to a georeference system can result in higher maintenance costs that stem from rerectifying images and resolving spatial database inaccuracies due to the placing of new features against a stretched and scaled base map.

Based on presentation and usage criteria, implementing a vector base map database is the most beneficial to the consortium. The vectorization of the base map data provides the analysts with the ability to query for base map data as well as pipeline data. The analyst can also perform proximity analysis by selecting base map features which reside within a specified distance from a pipeline (i.e. find all wildlife refuges within 20 km of the selected pipelines). Thus, the vectorization of the base maps increases the analysts' productivity by enhancing their analytical capabilities and access to up-to-date base maps and by reducing the maintenance costs associated with updating the base maps.

Data Storage

The base map database consists of all bodies of land between latitude 80N and 60S. While the remote facilities are only interested in data storage for areas which contain their pipelines, the central facility is interested in providing on-line access to the entire set of base maps. Table 1 provides information on the number and size of the base maps which the consortium plans to use as data sources for the base maps. The '% Coverage' column reflects the percentage of the Earth (between the latitudes of 80N and 60S) that is available in hardcopy maps sheets at a given scale.

Conversion of these hardcopy data sources to raster format requires a large-scale color scanner. The scanning process assigns color values for each pixel scanned. Each pixel's color value can be stored as an 8-bit (256 colors) or 24-bit (16.7 million colors) number. For this system, 8-bit storage is sufficient color representation for the base map database since the maps will be used primarily for locating pipeline features.

Map Scale	Number of Maps	% Coverage	Map Size (inch)	Map Area (in2)
1:1,000,000	320	100	18 x 18	324
1:250,000	5000	90	26.25 x 18	472.5
1:50,000	100,000	10	18.75 x 22	412.5

Table 1 Data Sources for the Base Map Database

An estimation of the amount of disk storage required to store a set of maps in raster format is calculated as

$$\textit{Disk Storage} = \textit{Area} \times \textit{Scan Resolution} \times \textit{Number of Maps} \times \textit{Number of Bytes per Pixel} \qquad \textit{Eqtn. (1)}$$

Table 2 shows the amount of disk storage required to store each of the map sets based on the scanner's resolution. Large scale scanners have scanning resolution which range from 200 to 500 pixels per inch or dots per inch (DPI).

The estimation of disk resources required to store a raster data set has been calculated based on the map size and the scanning resolution. Estimating the amount of disk resources required to store the various maps sets in vector format is more difficult. The size of a vector formatted map is based on the number of coordinate sets, number and size of the attributes which directly describe the features, and the precision at which the coordinates are stored. While the raster data sets always contains the same amount of pixels for the same size map, the vector data sets for the same maps depends on the density of features as well as each feature's shape. Estimations of the vector storage requirements are based on the size of existing vector data sets which have been digitized from the 1:1,000,000 scale map set. The size of the vector sets are based on half of the size of the raster sets which have been scanned at 200 dpi. Table 3 shows estimations for the vector storage of the three maps sets. Data storage costs where estimated at one dollar per megabyte.

Scanning Resolution ------->	200 DPI	200 DPI	300 DPI	300 DPI	400 DPI	400 DPI	500 DPI	500 DPI
Map Scale	**Single Map (GB)**	**Map Set (GB)**	**Single Map (GB)**	**Map Set (GB)**	**Single Map (GB)**	**Map Set (GB)**	**Single Map (GB)**	**Map Set (GB)**
1:1,000,000	0.013	4.15	0.0292	9.33	0.0518	16.6	0.081	25.9
1:250,000	0.0189	94.5	0.0425	213	0.0756	378	0.118	591
1:50,000	0.0165	10,500	0.0371	37,500	0.0660	66,000	0.103	103,000

Table 2 Raster Storage Requirements for Various Scanning Resolutions

Map Scale	Raster Storage at 200 DPI (GB)	Raster Costs	Vector Storage (GB)	Vector Costs
1:1,000,000	4.15	$4,150	2.3	$2,300
1:250,000	94.5	$94,500	47	$47,000
1:50,000	10,500	$10,500,000	5,500	$5,500,000

Table 3 Vector Storage Requirements for Three Map Sets

Scanning resolutions of 200 dpi are sufficient for the representation of base map features as long as the operator doesn't change the scale of the display to greater than 1:250,000. With a 200 dpi scan on a 1:250000 scale map, each pixel represents approximately 100 feet. Thus, added features will be located with an accuracy of +/- 100 feet. With increases in the scanning resolution, the raster image does not breakdown as fast as the analyst views the data at scales larger than the base map's scale. However, increasing the resolution results in increasing the data sizes which affect disk and network resources. Figure 28 provides a side-by-side comparison of raster and vector data storage requirements for the three map sets.

Storage Summary

Implementing the base maps in the raster format has a large impact on the amount of data storage required; the vector format provides for a reduced amount of storage. Creating the raster base maps with a 200 dpi resolution provides the analysts with sufficient accuracy and provides the administrators with a reduction in storage costs.

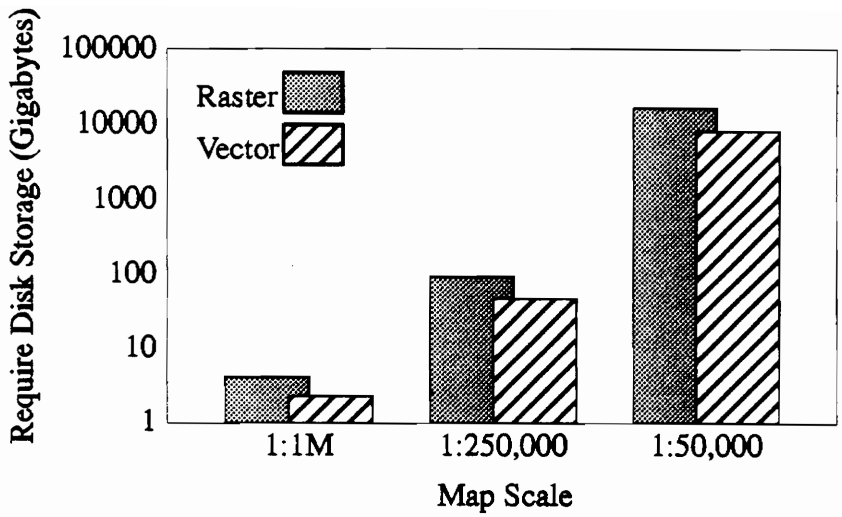


Figure 28 Storage Comparison (Vector Base Maps vs 200 DPI Raster Base Maps)

Display Performance

Although geographic information systems have been in existence for over twenty years, it is only in the last ten years that computer mapping systems have become a productive tool in the organization and display of spatial data due to hardware performance gains. This section quantifies the utilization of system resources and response times associated with the display of raster and vector data.

Since a majority of the consortium members utilize existing Sun Microsystem platforms, the performance tests were implemented on a prototype network of Sun equipment. The display performance trials were performed on Sun Microsystem's Sparcstation 10. The Sparc 10 uses a single 50 MHz RISC processor, 48 MB of memory, a 10Mb/sec SCSI (Small Computer Serial Interface) bus, with a 1.3 GB local SCSI disk drive. While the application software is accessed across the network, the test vector and raster data resides on the workstation's local disk.

Software monitoring tools, provided by Sun's operating system were utilized to collect data describing the utilization of the workstation's resources. The SunOS command 'vmstat' was used to collect data on the workstations utilization of disk and cpu resources. For all of the tests, the only applications running on the workstation were the operating system, the OpenWindows graphical user interface, the clock, the performance monitor, and Arc/Info. The display window in which the graphics were drawn was 8.5 by 11 inches.

A vector and raster version of a single 1:250,000 scale Defense Joint Operations Graphic (DJOG) was obtained for all of these performance tests. It has been assumed that this data set is the data used for meeting the system's performance parameters. The raster data set is 17.1 Megabytes. The vector data set is 2.5 Megabytes. The vector set consists of the following themes of data:

cul1_pt	[Culture I - Miscellanous Features]
cul2_pt	[Culture II - Miscellanous Features]
trns_pt	[Transportation]
trns_ln	[Transportation]
phys_ln	[Physiography]
hydr_ln	[Hydrography]
veg_py	[Vegetation]
cul2_py	[Culture II - Miscellanous Features]
trns_py	[Transportation]
phys_py	[Physiography]
cul1_py	[Culture I - Miscellanous Features]

Comparisons were made between the response times and percentage of CPU utilization for both data sets in Figure 29. The raster data utilized approximately a third of the CPU's processing power when compared to the vector data which required the entire CPU. The display of raster data is based on the CPU's ability to draw each pixel based on its colormap value. The display of vector data requires the CPU to calculate each feature's coordinates, to position each feature within the display window's coordinate system, and to display each feature in its assigned symbology.

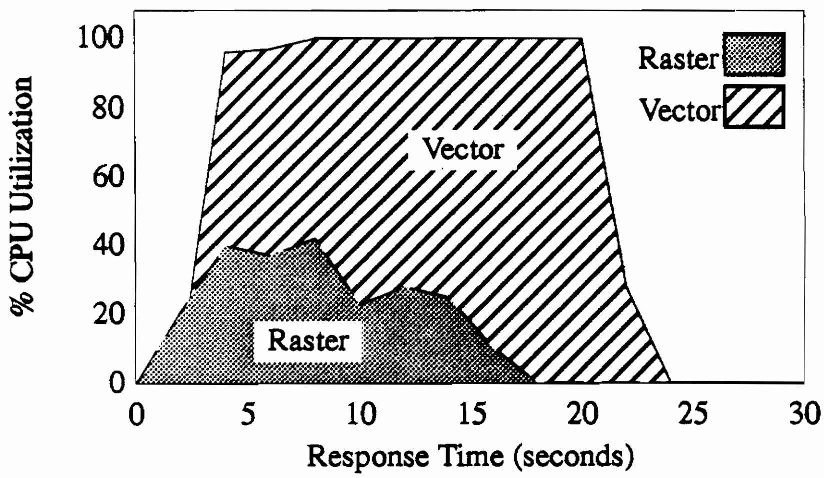


Figure 29   Response Time vs CPU Utilization (local data access)

Figure 30 shows the disk activity for both sets of data. With 17.1 megabytes of data, the raster format utilizes a significant amount of the local bus bandwidth when contrasted to the vector format.

With the large amount of feature data on the DJOG, the raster response time in presenting all the data is eight seconds faster than the vector response time. Since the vector set consisted of point, line, and polygon layers, further evaluation has been given to the response times associated with the different feature types within the vector data set. Figure 31 shows the response times associated with the three feature types. The 'trns_ln' layer accounts for the majority of time for the line features, and 'veg_py' layer accounts for almost the entire polygon section. Compared to line rendering, the display of polygons takes longer due to the need to fill and outline the polygon regions. The type of fill symbology also affects the response time.

Display Performance Summary

Despite the raster format's large size, displaying an image of the map results in a faster response time when contrasted to drawing all the vector layers. Drawing vector layers requires the CPU to calculate the various feature positions which takes more time than the raster. If the annotation of features is required, the raster image can present annotation information within the same time frame. The vector data requires additional draw time to place the annotation as well as to receive the information from the database. The main drawing advantage with vector data is the analyst's ability to display only the data of interest. Since analysts usually are only interested in a selected set of the data, it is possible that the analyst can receive faster response times when compared to viewing the entire raster image.

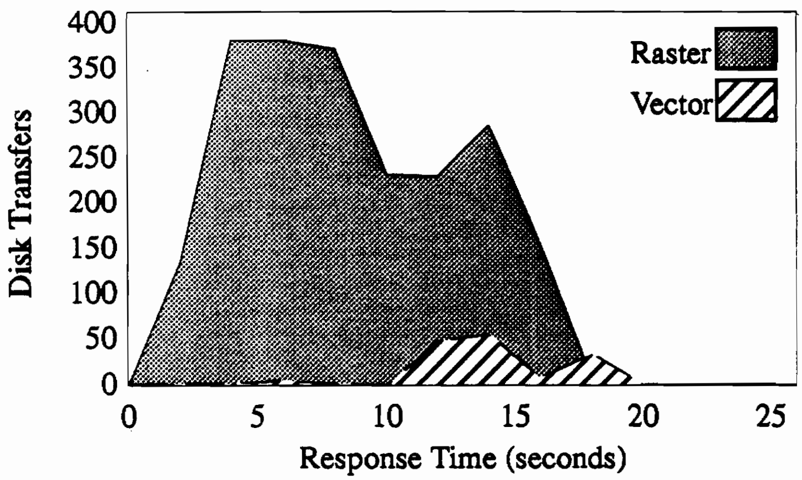


Figure 30 Response Time vs Disk Transfers (local data access)

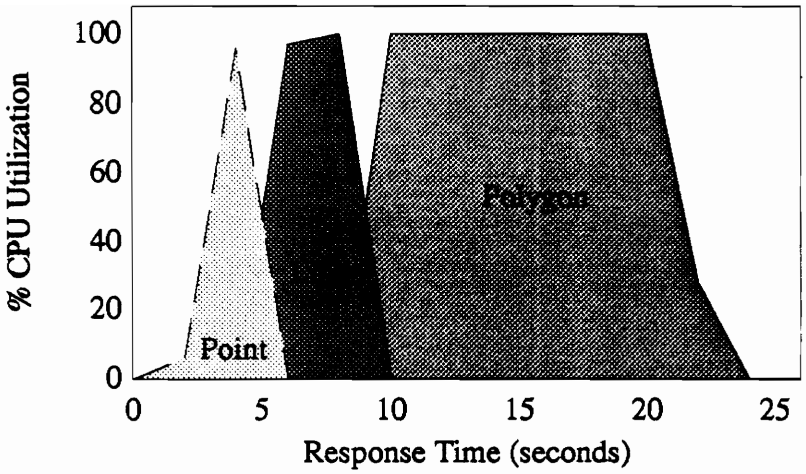


Figure 31 Response Time vs CPU Utilization (local data access) Vector Only

Impacts on the System Design

The processing power of the workstation's CPU is critical to the display's response time especially when working with vector data sets. The analyst workstation should be designed to support one user session at a time. Informal testing trials show a significant reduction in the applications and display response times if more than one analyst is utilizing the same CPU. Utilization of X servers from an existing PC to access the PTS interface through the server's CPU will result in slower performance for all applications running on the server unless extra CPUs are purchased and dedicated to X server access. The workstation's display performance meets the drawing performance requirements using a 50 Mz, 32-bit processor. Any existing system with slower processors should consider CPU upgrades. Memory expansions and fast local data buses are recommended, especially if raster data sets are used.

Network Performance

The network on which the system will be implemented consists of client workstations, similar to the Sparc10 used in the display testing, accessing a central server for applications and base map data across a 10 Mbit Ethernet local area network, LAN. The test data set consisted of a program which simulates the activity of three typical users sessions that consist of redrawing the display over an extent which covers sixteen DJOGs. The test data sets consisted of the vector and raster representations of the DJOGS.

As expected, due to the raster data set's size (each one approximately 17 MB), the transport of raster data sets across the network to assorted users results in rising collision rates, Figure 32, and slower response times, Figure 33. The vector data set's response time for eighteen user session remains the same despite a seven percent rise in the collision rate.

Network Performance Summary

The vector format is the optimum format for minimizing the amount of traffic on a local area network due to it smaller size.

Impacts on the System Design

Network segments provide better performance and support more users if vector base maps are utilized. With all the base map data residing on a centralized server and most users working against base map data, the network experiences some data traffic as well as application traffic. If raster data sets are utilized extensively, network designers should isolate the raster data users from other users so as not to slow down other LAN activities. If certain individuals or divisions operate on the same raster data sets, these data sets should be copied down to the local workstations or division servers in order to reduce the amount of network traffic. Finally, designers may research implementing a network which supports a larger bandwidth (more traffic).

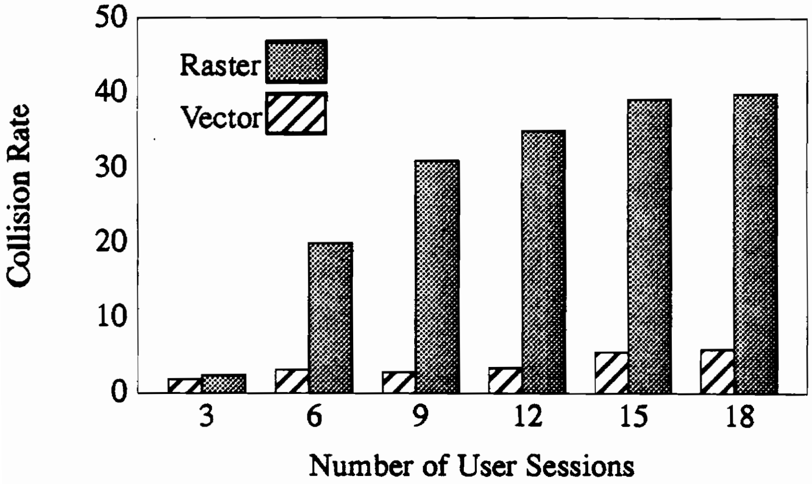


Figure 32 Number of User Sessions vs Collision Rate

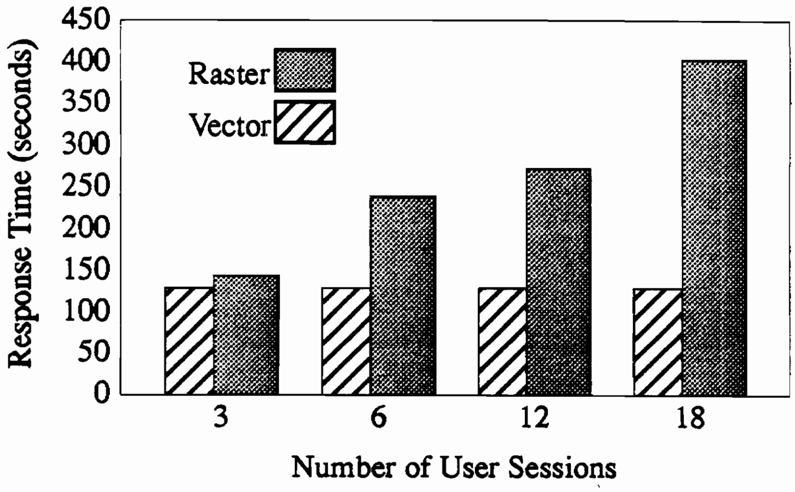


Figure 33   Number of User Sessions vs Response Time

Life-Cycle Costs Analysis for Base Map Implementations

The following section provides the life-cycle cost analysis for implementing each map set as either raster or vector base maps. The cost breakdown structure for the life-cycle costs associated with a vector base map database and a raster base map database is presented in Figure 34, Figure 35, and Figure 37. The structure's costs include research and design, map production, operational and maintenance costs associated with working with the base maps, and disposal costs. The cost structure is based on a seven year life-cycle which includes two years of development and five years of system operation (system's life span).

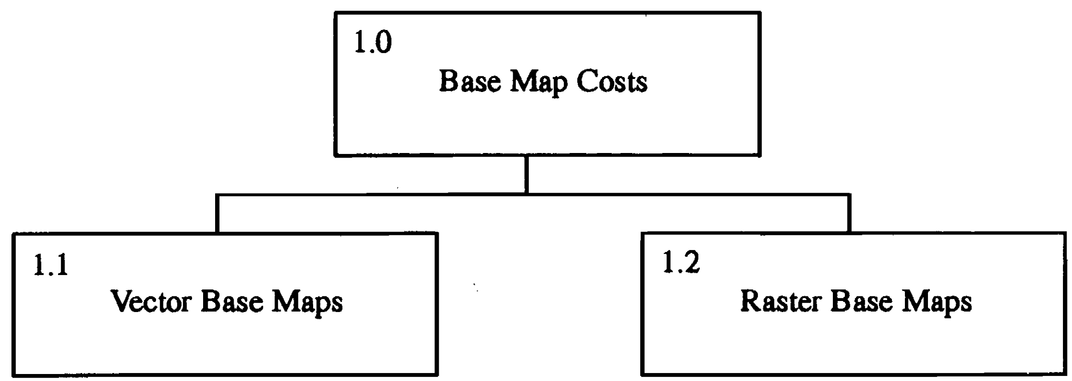


Figure 34   Base Map Database Cost Breakdown Structure

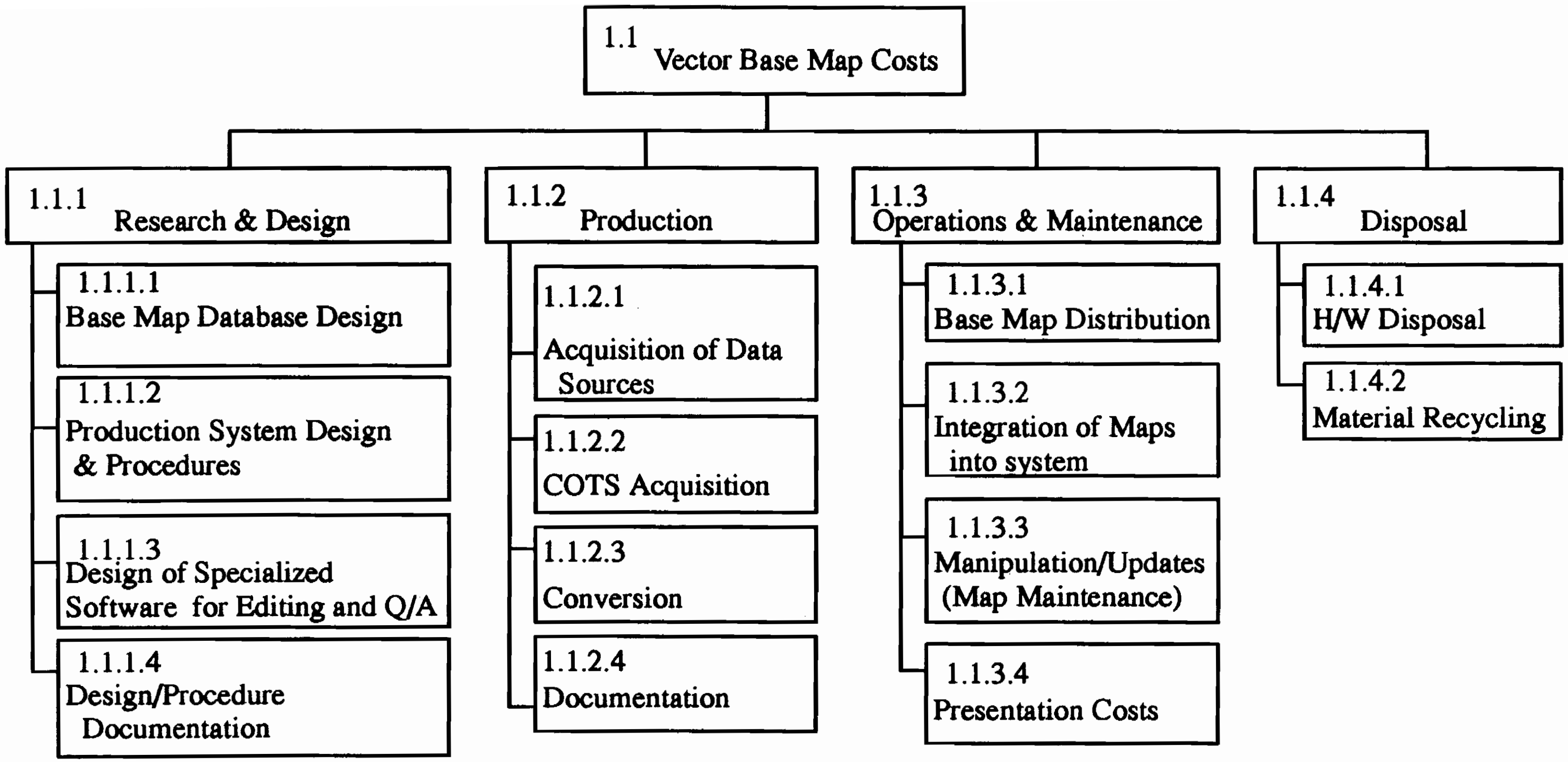


Figure 35 Vector Base Map Database Cost Breakdown Structure

Vector Base Map Database

The research and design cost item (1.1.1), Figure 35, consists of four cost items: database design (1.1.1.1), production system design (1.1.1.2), editing and quality assurance software design and development (1.1.1.3), and design documentation (1.1.1.4).

Storing base maps in vector format will require a large design effort in order to define the features necessary to assist the analysts with their data entry and analysis. For each set of maps, thematic layers of data must be defined such as transportation, petroleum, and drainage layers. Within each layer, individual features' various feature identification codes and names are defined. For example, one of the layers required by the analysts has a theme of 'cities'. Given a paper map, a conversion operator digitizes all of the cities on the paper map. The digitized layer may contain point features which give the centralized locations of the cities. If city limit borders are available, the operator may digitize the layer with polygon or line features which define the areas of the cities. Once the cities have been digitized, the individual city features require attribution such as names, populations, country, etc. For each layer of spatial data, categories of attribution codes are defined to describe the feature. Any feature coding (attribution) for a layer should be designed to support the analysts' needs to locate and identify particular feature types. The cost to design the database for a vector base map includes costs such as assessment of analysis needs, spatial database design (theme structures), and feature coding design. Special attention should be given to the assessment of the analysis need within the consortium; vectorization of the base maps will require a large investment of funds; vector coverages should only be made for features that will directly benefit the analysts' abilities to locate, identify, and analyze spatial data. The database design effort (1.1.1.1), Figure 35, will be performed by one individual over a six month period at a cost of $58,000.

With a large number of base maps requiring vectorization, detailed planning should also be given toward the method by which paper maps are converted to vector layers of data. The tasks within the production system include map scanning, raster to vector conversions, editing and attribution, quality assurance, and documentation. The consortium has been allocated a set amount of funds, $2,000,000/year, for the base map production. By defining and developing an efficient production system, the consortium hopes to maximize the production of high quality vector layers using the funds allocated to the consortium. The design of the production system (1.1.1.2), Figure 35, will be performed by one person over a three month period at a cost of $29,000.

Specialized software may be required to assist the operators in the clean-up of digitization errors that may result from the raster-to-vector conversion software and to assist in quality assurance (Q/A) procedures. To ensure the quality of a vector layer, Q/A procedures are established in which an operator compares digitized features to the original map features on the hardcopy map. Due to the large number of features to compare, statistical approaches are used against small areas of the maps. Specialized software programs can be written to perform automated checking of attribution and proper digitization. The costs of specialized software items in the cost breakdown structure includes the design, development, and test of these Q/A software modules as well as specialized spatial editing tools. The design and development of specialized software (1.1.1.3), Figure 35, will be performed by one person over two months at the cost of $19,000.

Documentation costs include the labor, production, and material costs for database design, data definition, production system design, operator training, and production system design documentation. The documentation costs (1.1.1.4), Figure 35, are projected to be $10,000 over a the two month period.

The production costs associated with the base maps (1.1.2), Figure 35, consist of four cost items: data source acquisition (1.1.2.1), COTS acquisition (1.1.2.2), conversion (1.1.2.3), and documentation (1.1.2.4).

The data source acquisition costs include all costs necessary to obtain and store a complete set of hardcopy maps at scales of 1:1000000, 1:250000, and 1:50000. These map sets have already been defined by the consortium's members. The acquisition cost (1.1.2.1), Figure 35, for all three map sets are projected to be a one time cost of $100,000.

The COTS acquisition costs include all additional hardware and software required to support base map production as well as any additional hardware costs that may be required to enhance the performance of the network. The vector conversion process will require 25 workstations with a single electrostatic plotter ($48,000) and large scale map scanner ($140,000) which will add up to a one-time COTS acquisition cost (1.1.2.2), Figure 35, of $700,000 in additional hardware and software.

The conversion costs include several subcost items: map scanning, digitization of features, attribution, quality assurance, and map product production. The map scanning item includes labor, training, and material costs necessary to produce raster color separates for input to the raster-to-vector conversion software. The digitization items include labor, software, and training costs required to extract vector features from a raster image. The attribution item includes labor and training costs required to populate the attribute fields associated with vector features. The quality assurance item includes labor and training costs required to ensure that each vector layer meets the spatial and attribution requirements. The product productions item includes labor and material costs associated with creating digital map products which may be distributed to remote facilities. The conversion costs are based on labor, overhead, and training costs associated with a staff of twenty-five people working over twelve months, $1,200,000 a year.

Preliminary estimates for the digitization of the required data, based on statistical feature density projections, require one man month of labor to digitize an entire map sheet from any of the 1:250000 and 1:50000 map sets. The 1:1000000 map set would only require the digitization of the petroleum layer since a commercial database (Digital Chart of the World) exists which would satisfy the analysts need for the other six themes. Therefore, man power estimates for digitizing the petroleum layer are half a man month per map. Table 4 shows the estimated time required to vectorize a set of base maps. Base map production begins in the second year of the system's life-cycle and ends in the seventh year. Given the life-cycle, the entire 1:1,000,000 scale map set and one third of the 1:250,000 scale map set could be vectorized. None of the 1:50,000 scale maps would be vectorized given the short system life-cycle.

The final cost item, documentation costs, includes data definition discrepancies and data quality remarks for individual vector layers. The documentation costs (1.1.2.4), Figure 35, for the production phase are project to be $10,000 a year.

Map Scale	Number of Available Maps	Conversion Period per Map (man months)	Total Conversion Time (man months)	Conversion Period w/ 25 people
1:1000000	320	.5	160	0.6 years
1:250000	4500	1	4500	15 years
1:50000	10,000	1	10000	33 years

Table 4 Estimated Man Months to Implement a Vector Database

The operation costs (1.1.3), Figure 35, associated with the base maps consist of four cost items: base map distribution (1.1.3.1), base map integration (1.1.3.2), manipulation/update (1.1.3.3), and presentation (1.1.3.4).

The base map distribution item, $15,000/year per facility, includes labor, storage, material, and shipping costs required to distribute finished base map products to remote facilities. The integration items, $20,000/year per facility, includes the labor costs required to install the vector base maps on the system. The manipulation/update costs are the maintenance costs associated with modifying the vector database to record changes (new buildings, modified country borders, etc.) in the location of spatial features. The maintenance costs are expected to be performed by one person at a cost of $40,000 a year. The presentation costs are expect to result in productivity and analysis gains at each facility estimated at earn $250,000 a year for each map set beginning in the system's second year. The negative values for the presentation costs represent assumed financial gains to the consortium due to the productivity and analysis benefits that a vector base map library will provide.

The disposal costs (1.1.4), Figure 35, associated with the base maps consists of two cost items: hardware disposal (1.1.4.1) and recycling (1.1.4.2). The hardware disposal item includes, $5,000, the cost to remove, dismantle, and dispose of non-recyclable pieces of equipment at all the facilities. The recycling item, $1000, includes the costs to transport and recycle paper maps, magnetic media, and other recyclable items at all the facilities.

Table 5 presents the present equivalent cost of each item using an interest rate of 6%. Figure 36 summarizes the vector base map costs associated with each cost item.

Vector Cost Item	Estimated Cost	Year(s) Cost Occurs	Present Value Cost
1.1.1.1 Base Map Design	$58,000	2	$51,620
1.1.1.2 Production System Design	$29,000	2	$25,810
1.1.1.3 Specialized S/W	$19,000	2	$16,910
1.1.1.4 Design Documentation	$10,000	2	$8,900
1.1.2.1 Data Source Acquisition	$100,000	2	$89,000
1.1.2.2 COTS Acquisition	$700,000	2	$623,000
1.1.2.3 Conversion	$600,000 $1,200,000/year	2 3-7	$534,000 $4,498,920
1.1.2.4 Production Documentation	$5000 $10,000/year	2 3-7	$4,450 $37,491
1.1.3.1 Base Map Distribution	$15,000/year	3-7	$56,236
1.1.3.2 Base Map Integration	$20,000/year	3-7	$74,982
1.1.3.3 Manipulation/ Update	$40,000/year	3-7	$149,964
1.1.3.4 Presentation Costs	-$200,000	3-7	-$749,820
1.1.4.1 Disposal Fees	$5000	7	$3,325
1.1.4.2 Recycling	$1000	7	$665
Total Cost			$5,425,353

Table 5 Total Vector Base Map Costs

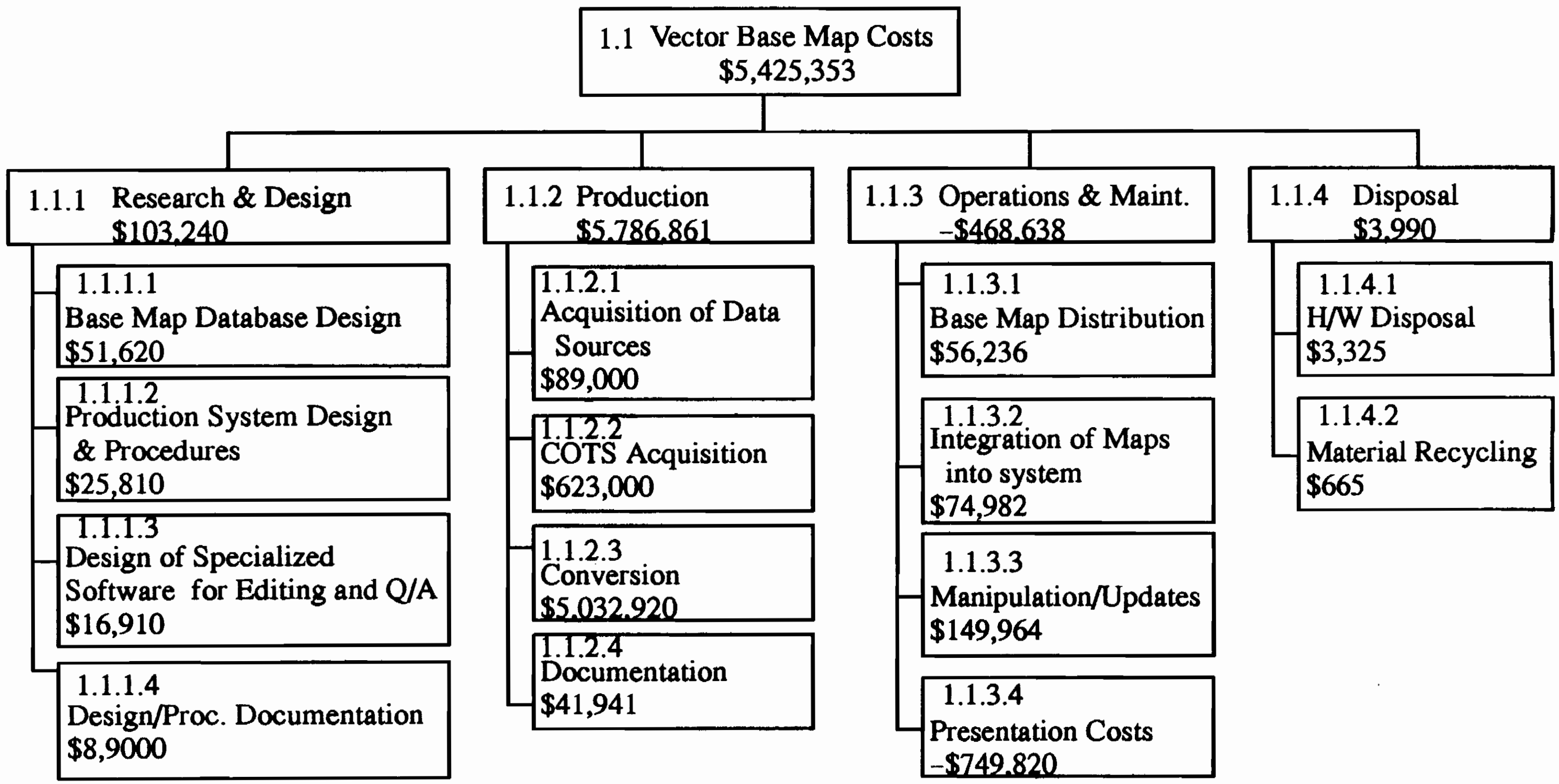


Figure 36 Vector Base Map Database Costs

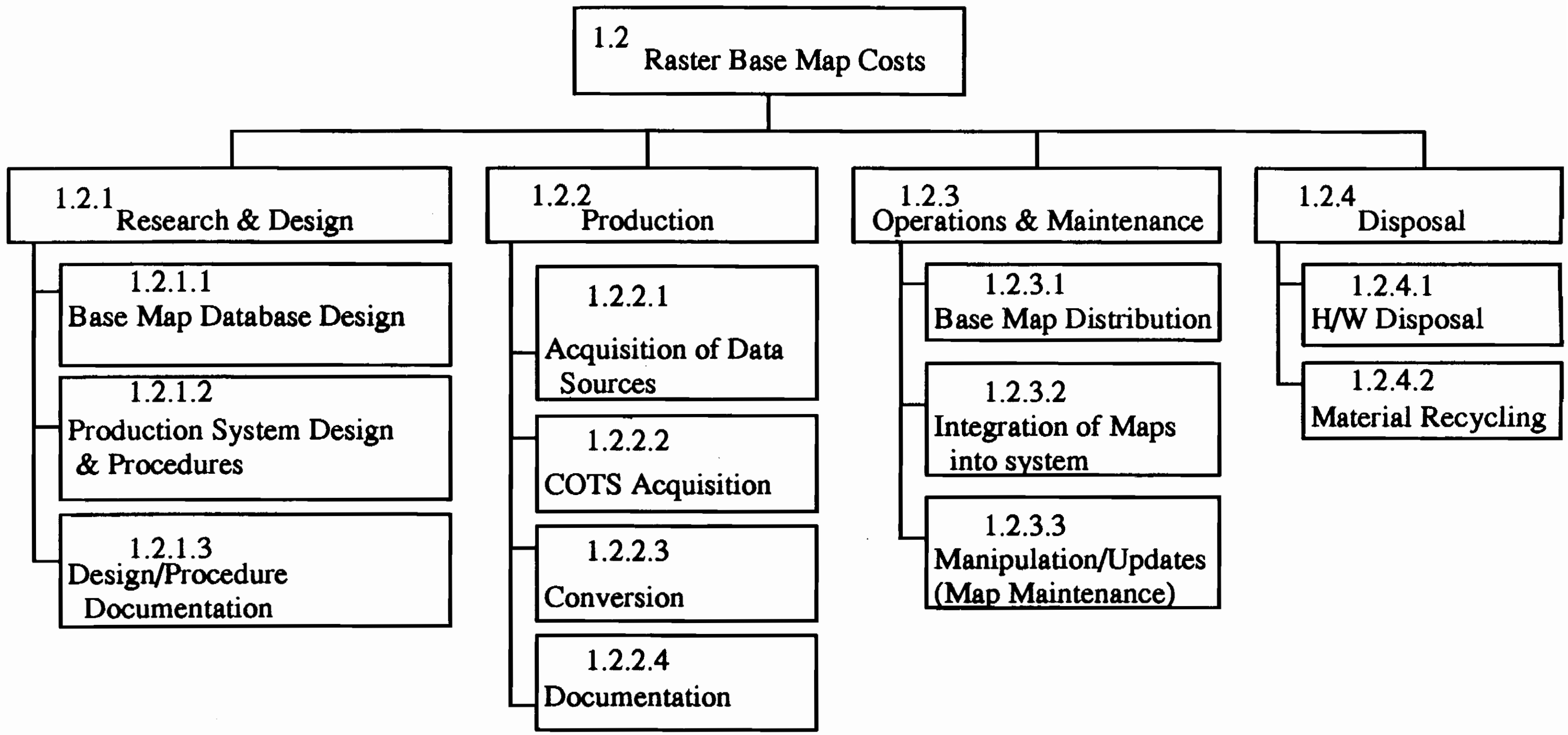


Figure 37 Raster Base Map Database Cost Breakdown Structure

Raster Base Map Database

The research and design cost item (1.2.1), Figure 37, consists of three cost items: database design (1.2.1.1), production system design (1.2.1.2), and design documentation (1.2.1.3). There are no specialized software requirements to assist in the Q/A procedures.

Storage of the base maps in raster format will require a smaller effort when compared to a vector implementation. The attribute database will only consist of attributes describing the entire map sheet. Analysts and the interface software may retrieve base maps based on searches against the map's description. Other design issues include resolution and registration requirements, priority scheduling, as well as allocation to storage devices. The database design effort (1.2.1.1), Figure 37, will be performed by one individual over a four month period at a cost of $25,000.

With a large number of base maps requiring rasterization, detailed planning should be given toward the method by which hardcopy maps will be scanned and registered to a coordinate system. The tasks within the production system include map scanning, image editing and registration, attribution, quality assurance, documentation, and distribution. By defining and developing an efficient production system, the consortium hopes to maximize the production of high quality registered images using using the funds allocated to the consortium. The design of the production system (1.2.1.2), Figure 37, will be performed by one person over a two month period at a cost of $12,000.

Documentation costs include the writing, production, and material costs for database design, data definition, production system design, operator training, and production system design documentation. The documentation costs (1.2.1.3), Figure 37, are projected to be $15,000.

The production costs (1.2.2), Figure 37, associated with the base maps consist of four cost items: data source acquisition (1.2.2.1), COTS acquisition (1.2.2.2), conversion (1.2.2.3), and documentation (1.2.2.4).

The data source acquisition costs include all costs necessary to obtain and store a complete set of hardcopy maps at scales of 1:1000000, 1:250000, and 1:50000. These map sets have already been defined by the consortium's members. The acquisition cost (1.2.2.1), Figure 37, for all three map sets are projected to be $100,000.

The COTS acquisition costs include all additional hardware and software required to support base map production as well as any additional hardware costs that may be required to enhance the performance of the network. The raster conversion process will require 25 workstations with a single electrostatic plotter ($48,000) and three large scale map scanner ($420,000) which will add up to a one-time COTS acquisition cost (1.2.2.2), Figure 37, of $980,000 in additional hardware and software.

The conversion costs include several subcost items: map scanning, image editing and registration, attribution, quality assurance, and map product production. The map scanning item includes labor, hardware and software, training, and material costs necessary to produce color rasters. The editing and registration item includes labor, software, and training costs required to adjust the image's properties and to assign coordinate information. The attribution item includes labor and training costs required to populate the attribute fields associated with describing the individual map sheet (map scale, index number, extent, scan date, etc.). The quality assurance item includes labor and training costs required to ensure that each raster map meets the spatial and attribution requirements. The product productions item includes labor and material costs associated with creating digital map products which are distributed to remote facilities. The conversion costs are based on labor, overhead, and training costs associated with a staff of twenty-five people working over twelve months, $1,200,000 a year.

Preliminary estimates, based on statistical feature density projections, require a quarter man month of labor to rasterize an entire map sheet from any of the three map sets. Assuming the consortium's facility can support a conversion team of 25 people, the conversion staff can produce 1200 raster map products in a year. Table 6 shows the total estimated time period required to rasterize a set of base maps. Base map production begins in the second year of the system's life cycle and ends in the seventh year. Given the life cycle, all of the 1:1,000,000 scale and 1:250,000 scale maps could be rasterized. In addition, approximately one eighth of the 1:50,000 scale maps could be rasterized before the end of the life cycle.

The documentation costs include data definition discrepancies and data quality remarks for individual map sheets. The documentation costs for the production phase (1.2.2.4), Figure 37, are project to be $5,000 for each year.

Map Scale	Number of Available Maps	Conversion Period per Map (man months)	Total Conversion Time (man months)	Conversion Period w/ 25 people
1:1000000	320	.25	80	0.5 years
1:250000	4500	.25	1125	3.8 years
1:50000	10,000	.25	2500	8.3 years

Table 6 Conversion Costs to Implement a Raster Database

The operation costs (1.2.3), Figure 37, associated with the base maps consist of three cost items: base map distribution (1.2.3.1), base map integration (1.2.3.2), and manipulation/update (1.2.3.3).

The base map distribution, $15,000/year per facility, item includes labor, storage, material, and shipping costs required to distribute finished base map products to remote facilities. The integration items, $10,000/year per facility, includes the labor costs required to install the raster base maps on the system. The manipulation/update costs are the maintenance costs associated with modifying the raster database to record changes (new buildings, modified country borders, etc.) in the location of spatial features. The maintenance costs are expected to be performed cost of $100,000/year for the reregistration of maps at the various remote facilities as well as for the occasional rescanning of new editions. The raster implementation of the base maps will not result in any productivity and analysis gains due to presentation.

The disposal costs (1.2.4), Figure 37, associated with the base maps consist of two cost items: hardware disposal (1.2.4.1) and recycling (1.2.4.2). The hardware disposal item, one-time fee of $5,000 for all sites, includes the cost to remove, dismantle, and dispose of non-repairable or retired pieces of equipment at all sites. The recycling item, one-time fee of $1000 for all sites, includes the costs to recycle paper maps, magnetic media, and other recyclable items at all sites.

Table 7 presents the present equivalent cost of each item using an interest rate of 6%. Figure 38 summarizes the raster base map costs associated with each cost item.

Raster Cost Item	System Cost	Year(s) Cost Occurs	Present Value Cost
1.1.1 Base Map Design	$25,000	2	$22,250
1.1.2 Production System Design	$12,000	2	$10,680
1.1.3 Design Documentation	$15,000	2	$13,350
1.2.1 Data Source Acquisition	$100,000	2	$89,000
1.2.2 COTS Acquisition	$980,000	2	$872,200
1.2.3 Conversion	$600,000 $1,200,000/year	2 3–7	$534,000 $4,498,920
1.2.4 Documentation	$2,500 $5,000/year	2 3–7	$2,225 $18,745
1.3.1 Base Map Distribution	$15,000/year	3–7	$56,236
1.3.2 Base Map Integration	$10,000/year	3–7	$37,491
1.3.3 Manipulation/Update	$100,000/year	3–7	$374,910
1.4.1 Disposal Fees	$5000	7	$3,325
1.4.2 Recycling	$1000	7	$665
Total Cost			$6,533,897

Table 7 Total Raster Base Map Costs

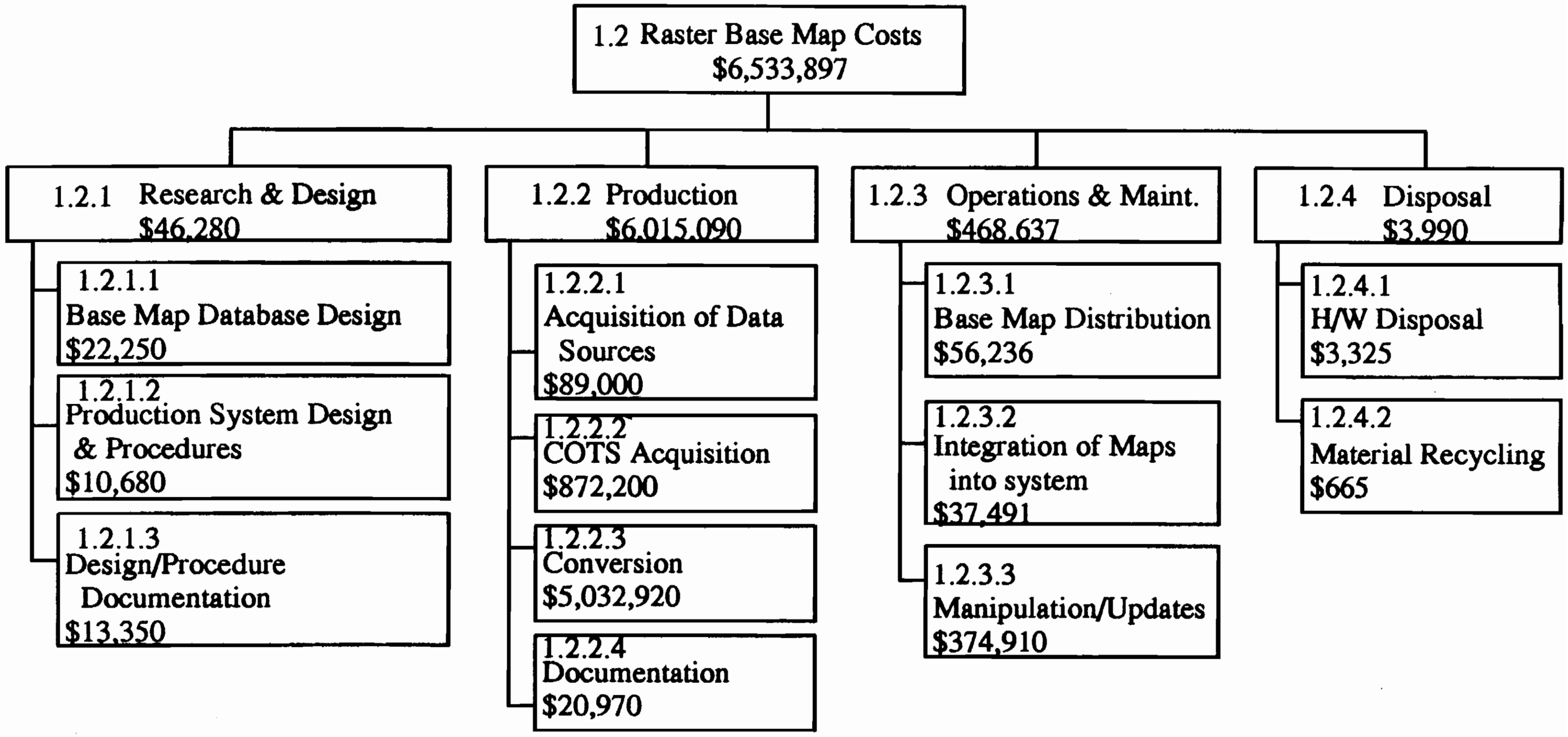


Figure 38 Raster Base Map Database Costs

Base Map Life-Cycle Cost Summary

Table 8 shows the total, present value costs associated with a raster and a vector base map database using an interest rate of 6% over a seven year life-cycle. Based on the availability of funds ($14,000,000) and a seven year life-cycle, the consortium would only have 1/3 of the 1:250:000 scale maps vectorized and none of the 1:50,000 scale map sets vectorized. For an extra 1.1 million dollars, the entire 1:250,000 scale map set and 1/8 of the 1:50,000 scale map set can be rasterized. Since the 1:1000000 scale product Digital Chart of the World, DCW, provides information of all the required data themes with the possible exception of certain petroleum features, the DCW product should be utilized for the 1:1000000 map set in addition to having the petroleum coverages digitized at the 1:1000000 scale. Over the system's life-cycle, the conversion operators should first focus on creating a raster base map library of the 1:250:000. Since the 1:50,000 map set only covers 10% of the world, this map set should be rasterized following the complete rasterization of the 1:250,000 scale maps.

Implementation	Total Base Map Cost
Vector	$5,425,353
Raster	$6,533,897

Table 8 Life-Cycle Cost Comparison for Base Map Formats

Base Map Analysis Conclusions

Without cost and schedule restraints, the vector format provides for the best implementation of the PTS base map database for all three sets of maps. The vector format provides a flexible and more accurate representation of the base maps. Analysts may control the data they are viewing and may locate features within the base maps to assist in their analysis. However, given the amount of resources available to support the base map production over the term of the system's life-cycle, the production costs and time required to digitize the entire world at scales of 1:250000 and 1:50000 are not feasible to the consortium.

For the 1:1000000 scale maps, it is recommended that the commercial product, Digital Chart of the World (DCW) be obtained for each facility. This vector product will provide a sufficient amount of spatial and attribute data on all of the consortium's required data themes. This product will provide 100% coverage for all areas of interest. This vector database has a limited amount of petroleum features. An evaluation should be made to determine if the petroleum features within DCW are sufficient for the consortium's analysis needs or if further digitization will be required.

For the 1:250000 scale maps, it is recommended that the conversion staff begin design and production activities to implement a raster base map database using these maps.

Since the time required to produce a raster library at 1:50000 extends beyond the system's life-cycle, it is recommended that the individual map sheets be scanned and registered on a map by map basis at the special request of individual companies while the 1:250,000 scale maps are in production. Since the 1:50000 map set only covers 10% of the area of interest and requires a prohibitive amount of resources to be digitized, the base map database should not include 1:50000 scale maps until the 1:250000 maps are all rasterized.

Preliminary System Configurations

Analysts, who are entering pipelines and evaluating spatial locations, can be expected to operate on base maps which provide the greatest degree of accuracy. Based on previous analysis, the analysts will be utilizing the raster base map library at a scale of 1:250000. The utilization of raster data across the network generates a significant amount of traffic as more and more analysts access the base map database. The network design should focus on implementing LANs which provide large bandwidths to increase traffic flow and shorter segments (assuming an Ethernet protocol) to reduce the amount of time data packets are on the network thereby reducing the collision rates. A LAN segment can support approximately 12 active users without dramatic reductions in network and display performance. Routers should be utilized at the network hubs to isolate user groups, such as the conversion operators at the central facility, who will work primarily on raster data.

To reduce the amount of raster data sent out across the LAN, raster data which is used repeatedly can be copied down to the analyst's workstation or located at a group/division level server. This configuration is shown in Figure 40. In all of the configurations, each workstation should have an addition 1 GB disk capacity and at least 16 MB of additional memory. This will provide disk space for copied down rasters and provide for faster redisplays of the raster data.

Figure 39 through Figure 41 present the preliminary hardware configurations based on the base map analysis.

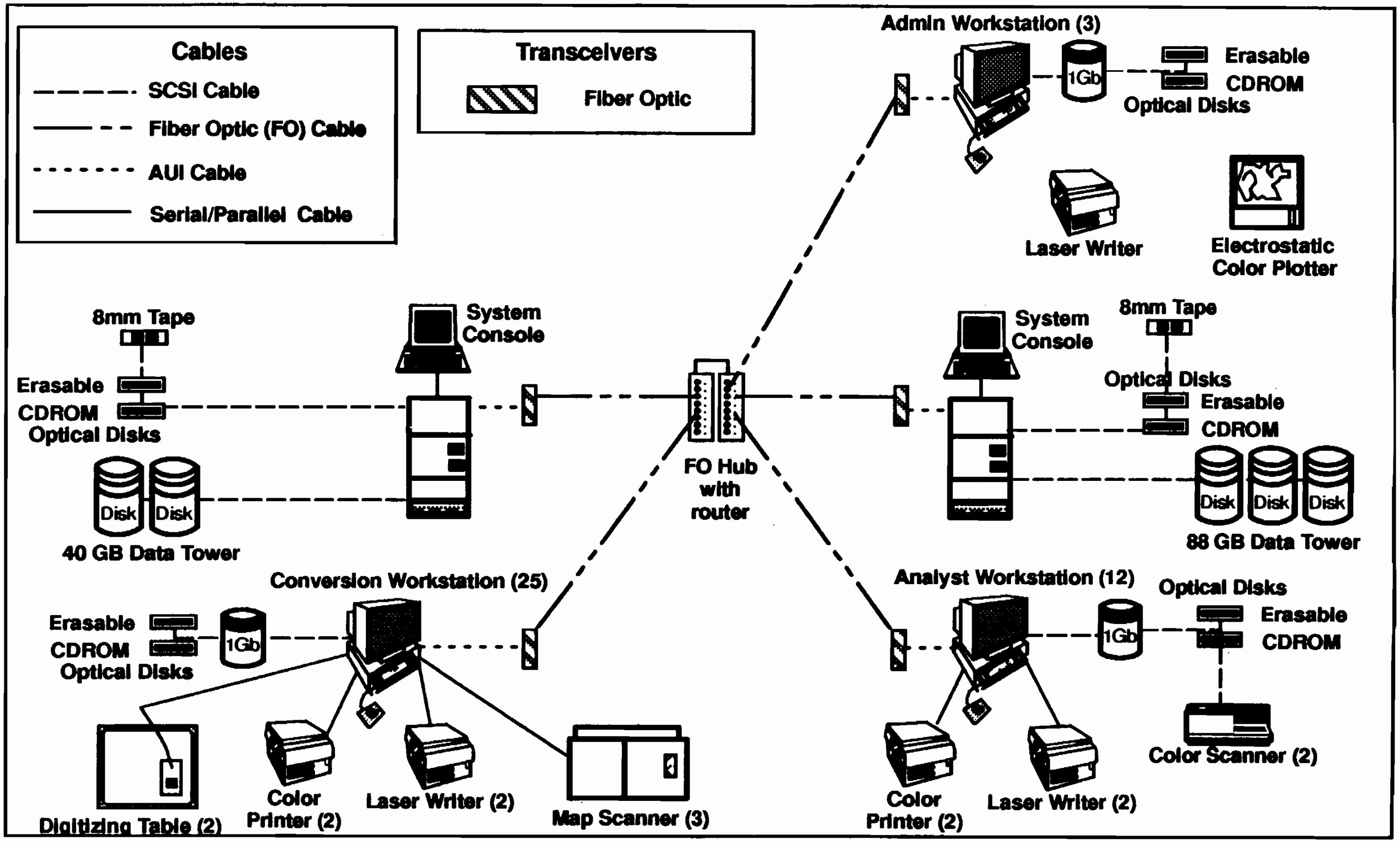


Figure 39 Central Facility's Preliminary Hardware Configuration

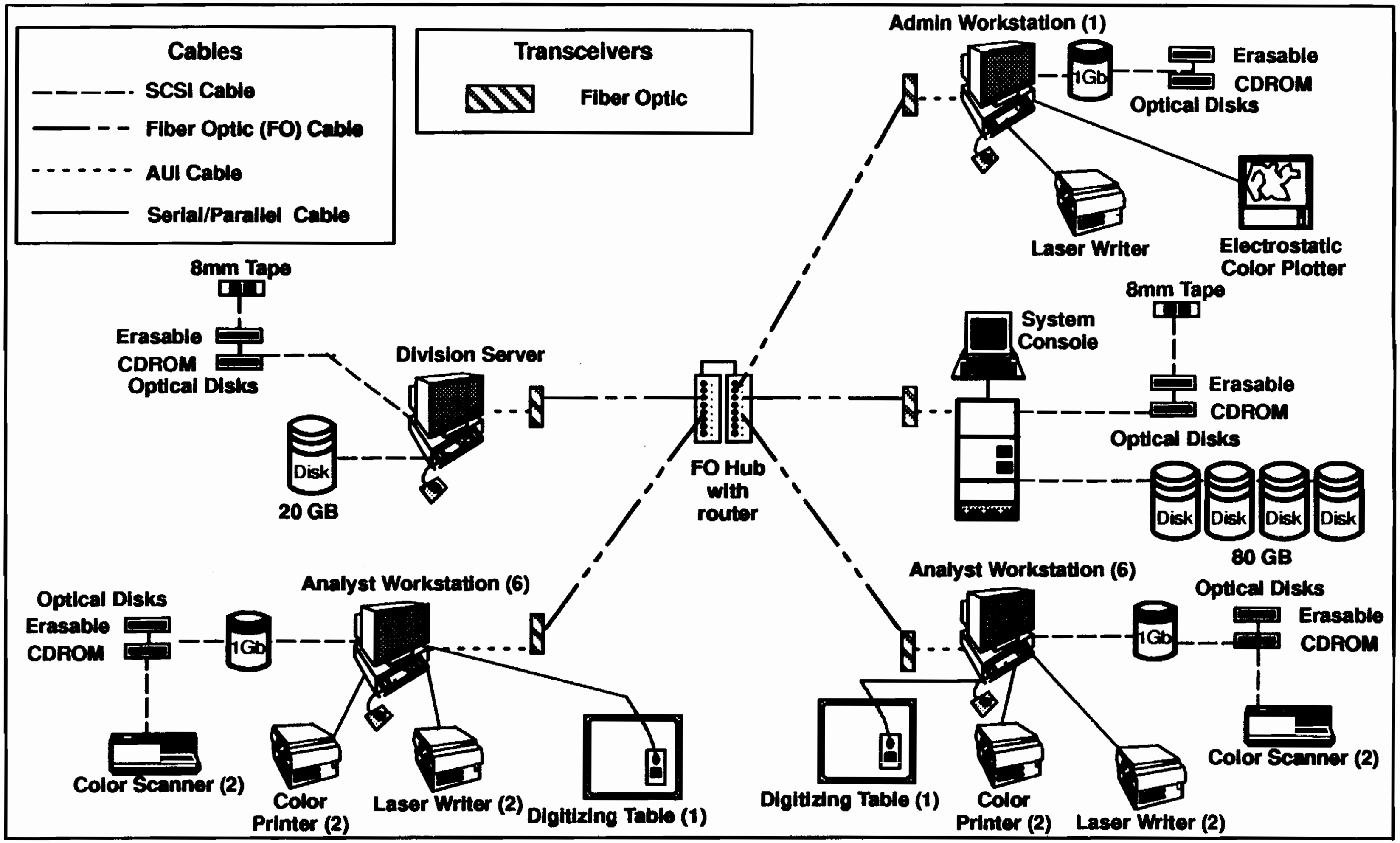


Figure 40 Large Remote Facility Preliminary Hardware Configuration

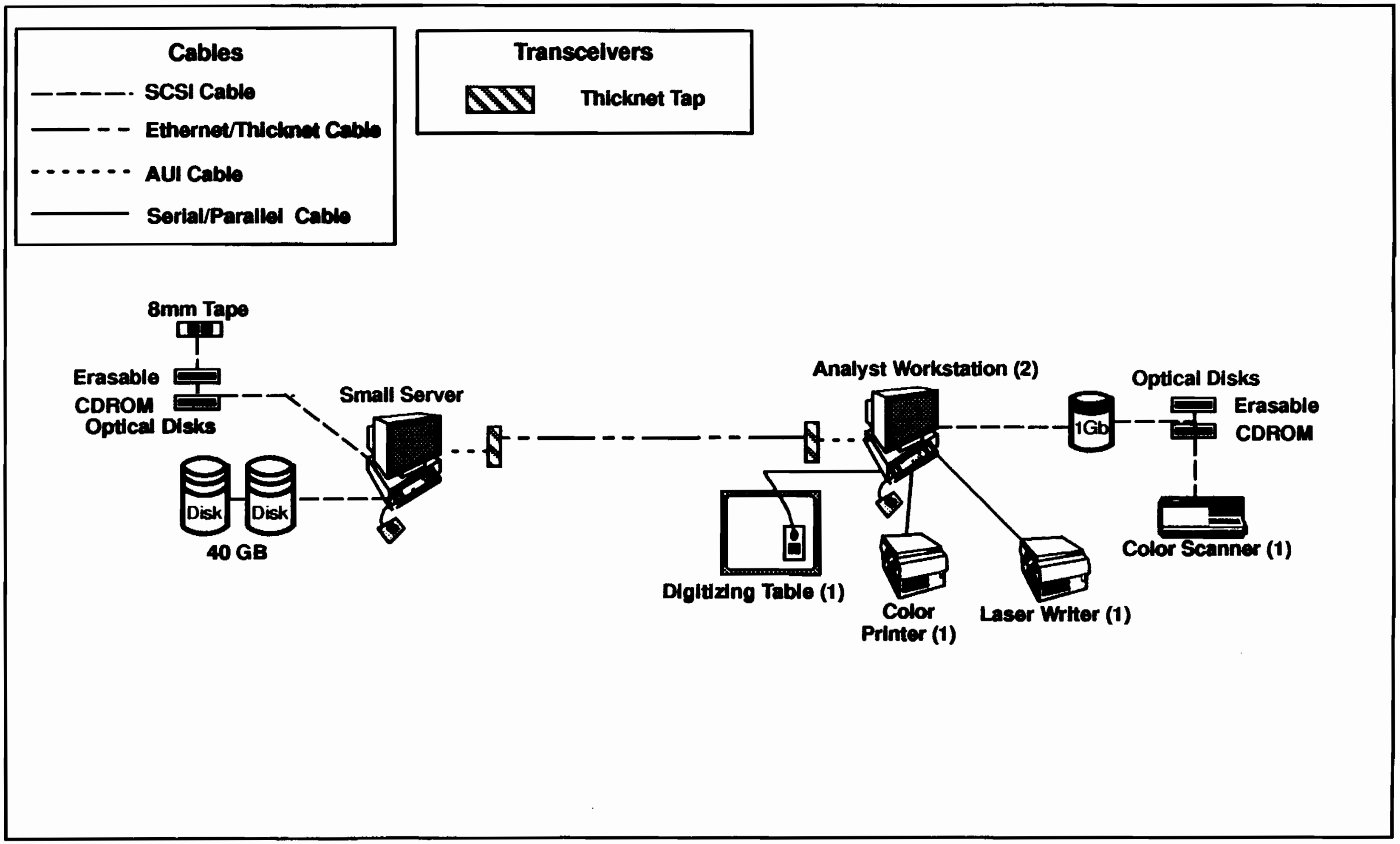


Figure 41 Small Remote Facility Preliminary Hardware Configuration

Further Recommended Studies

The decision to represent the majority of the base maps in a raster format is primarily based on the consortium's cost and schedule restraints. For the consortium's data entry and display needs, the raster format provides analysts with sufficient locational information to assist in their placement of pipeline features. In addition, the consortium can have its entire region of interest rasterized within the system's life span. In contrast, the cost and time required to digitize the base maps is difficult to justify unless the benefits of digitizing the map sets can be better quantified.

Additional research that focuses on quantifying the benefits of a vector base map database may produce changes in the base map cost analysis which may favor selection of the vector format. While the system's life span is approximately five years, the data's life span will be much longer. Data created within the PTS system will most likely be downloaded or converted to data for future generations of systems. By extending the life span on the data and by identifying and quantifying additional analysis capabilities that are available when utilizing vector base maps, the benefits of a vector base map database may outweigh the large costs of creating the vector database.

ACRONYMS

ARB	Analysis Review Board
COTS	Commercial Off The Shelf
DCW	Digital Chart of the World
DJOG	Defense Joint Operations Graphic
DPI	Dots per Inch
FO	Fiber Optic
GB	Gigabyte
GIS	Geographic Information System
GPS	GeoPositioning Satellite
LAN	Local Area Network
MB	Megabyte
PTS	Pipeline Tracking System
SCSI	Small Computer Serial Interface
TBD	To Be Determined

REFERENCES

1) Blanchard, Benjamin S. and Wolter Fabrycky. Systems Engineering and Analysis. Prentice Hall: New Jersey. 1990

2) Environmental Systems Research Institute. ARC/INFO Data Model, Concepts, and Key Terms. ESRI: Redlands March 1992

3) Environmental Systems Research Institute. Image Integration ESRI: Redlands March 1992

4) Environmental Systems Research Institute. Map Projections and Coordinate Management. ESRI: Redlands March 1992

5) Tomlin, C. Dana. Geographic Information Systems and Cartographic Modeling. Prentice Hall: New Jersey 1990

6) Sun Microsystems, Inc. Advanced System Administration. Sun Microsystems, Inc.: Ohio 1990

ENDNOTES

(1) System Engineering and Analysis pg. 20–21

(2) Ibid., pg. 35–36

(3) ARC/INFO Data Model, Concepts, and Key Terms pg. 1–8

(4) Ibid., pg. 1–6