

**INTEGRATION OF CONVENTIONAL WEB-BASED GIS AND
GLOBALBASE USING GEOMETRIC TRANSFORMATIONS AND
HIGH-LEVEL MIDDLEWARE**

MAULANA ABDUL AZIZ

UNIVERSITI SAINS MALAYSIA

2008

**INTEGRATION OF CONVENTIONAL WEB-BASED GIS AND
GLOBALBASE USING GEOMETRIC TRANSFORMATIONS AND
HIGH-LEVEL MIDDLEWARE**

by

MAULANA ABDUL AZIZ

**Thesis submitted in fulfillment of the
requirements for the degree
of Master of Science**

JUNE 2008

ACKNOWLEDGEMENTS

All the praises and my greatest gratitude to Allah SWT, the Grandest and Almighty, Most Gracious, Most Merciful, and Holder of all knowledge for giving me the chance, time, and strength to perform this study and for all the chances He has given to me until now. My greatest gratitude to the Master of the Messengers, Muhammad SAW, for the teachings and love that he has spread to the whole world.

This work has its roots in the teaching, help, patience, inspiration, and support of a great number of people, to whom I wish to express my heartfelt gratitude.

Simply put, Associate Professor Dr. Rahmat Budiarto is the reason I embarked on this study; and the reason I had faith that this work, with sufficient diligence, yield a host of new insights. His inspiration, guidance, patience, and invaluable support have provided me with the confidence to continue the work until the end. The energy and respect that he affords to both his students' efforts, and to his students themselves, has made it a privilege to be supervised by him.

Likewise, I am also grateful to my second supervisor, Associate Professor Dr. Abdullah Zawawi Hj. Talib, for his helps and comments during some events. In addition, Professor Dr. Rosni Abdullah as the Dean of School of Computer Sciences, Universiti Sains Malaysia, and Associate Professor Dr. Sureswaran Ramadass as the Director of NaV6 for providing space and facilities to conduct this research work.

Special thanks go to Associate Professor Hirohisa Mori from International Research Center for Japanese Studies, Japan and Associate Professor Dr. Jamaluddin Mohd. Ali from School of Mathematical Sciences, Universiti Sains Malaysia for their fruitful discussions and generosity in sharing their knowledge. I am also gratitude for Professor Haruhiro Fujita

along with his team from Toyo University, Japan for their assist and kindness during GLOBALBASE training sessions.

I would also like to thank Muhammad Fermi Pasha and Ibrahim Umar for readily providing suggestions and expert opinions on my study in numerous discussions. Also thanks to Eric Ho Yew Leong, Ardiansyah, Shoumananda Rangkuti, and Hapsari Puspitaloka for their help and assistance in initial preparations for the study and during thesis writing. Moreover, my heartfelt thanks to all my friends and colleagues, both from inside and outside of Universiti Sains Malaysia, for their supports and make my life more colorful and a well spent during my postgraduate studies.

Last but not least, a special thank you to my parents, Aijub Muchtar and Faridawaty Lelo, both my sisters, Khuria Amila and Ursula Salamah, and my grandmother, Azimah Tamin, whose love, faith, and constant support fuel all of my endeavors. I dedicate this thesis to my parents as their late silver anniversary gift. Without their support and understanding, this work would never reach its end. May Allah SWT grant you the best reward for all of your kindnesses, assistances, and supports.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF ABBREVIATION	xi
LIST OF APPENDICES	xii
ABSTRAK	xiii
ABSTRACT	xv

CHAPTER ONE : INTRODUCTION

1.1	Background	1
1.2	Problem Statement	3
1.3	Research Objectives	3
1.4	Contribution of This Thesis	4
1.5	Thesis Outline	5

CHAPTER TWO : LITERATURE REVIEW

2.1	Definition of Geographic Information System	7
2.2	Spatial Data Model in GIS	8
	2.2.1 Raster Data Model	9
	2.2.2 Vector Data Model	10
	2.2.3 Metadata in GIS	12
2.3	Geodesy Concepts	14
2.4	Architecture of GIS	15
	2.4.1 Client-Server GIS	17
	2.4.2 Internet GIS	18
	2.4.3 Mobile GIS	20
2.5	GLOBALBASE	22
	2.5.1 GLOBALBASE Architecture	22
	2.5.2 GLOBALBASE Data Structure	24

2.5.3	GLOBALBASE Protocol	26
2.6	Summary	28

CHAPTER THREE : METHODOLOGY AND DESIGN

3.1	Spatial Attributes Transformation for Interoperability of Various Spatial Data	30
3.1.1	Coordinate Conversion between Coordinate Systems	31
3.1.1(a)	Linear Transformation Based on Two Pairs of Control Points	32
3.1.1(b)	Linear Transformation Based on Three Pairs of Control Points	34
3.1.2	Transformation on LoD	37
3.1.3	Calculating Resolution of a Location on a Map	42
3.2	Towards Designing a High-Level Middleware	45
3.2.1	Google Maps API	48
3.2.2	Receive Handler	49
3.2.3	LoD Converter	51
3.3	Extended Cosmos Overall Design	51
3.4	Summary	57

CHAPTER FOUR : IMPLEMENTATION DETAILS

4.1	Transformation Process between Spatial Attributes in GLOBALBASE and Google Maps	59
4.1.1	Spatial Attributes in GLOBALBASE	60
4.1.2	Spatial Attributes in Google Maps	63
4.1.3	Spatial Attributes Transformation from GLOBALBASE to Google Maps	66
4.1.4	Spatial Attributes Transformation from Google Maps to GLOBALBASE	72
4.2	Cosmos GUI Integration	75
4.3	Summary	77

CHAPTER FIVE : EXPERIMENTATION AND RESULT ANALYSIS

5.1	Experimentation Direction	79
5.1.1	Experimentation on Rational Cubic Spline Interpolation	80
5.1.2	Experimentation on Spatial-Based Metadata Transformation between GLOBALBASE and Google Maps	81
5.1.2(a)	Server Installation Setup	82
5.1.2(b)	Spatial Data Setup	84
5.1.2(c)	Experimentation Procedure	90
5.2	Result and Analysis	93
5.2.1	Result and Analysis of the First Experimentation	93
5.2.2	Result and Analysis of the Second Experimentation	96
5.2.2(a)	Coordinate Transformation Result	96
5.2.2(b)	LoD Transformation Result	106
5.2.2(c)	Other Experimentation Results	109
5.3	Summary	115

CHAPTER SIX : SUMMARY AND FUTURE WORKS

6.1	Revisiting Research Contributions	117
6.2	Summary of Thesis	118
6.3	Future Works	120

BIBLIOGRAPHY	121
---------------------	-----

APPENDICES

Appendix A [High-Level Middleware (Part 1)]

Appendix B [High-Level Middleware (Part 2)]

Appendix C [High-Level Middleware (Part 3)]

LIST OF TABLES

	Page	
2.1	Characteristics of several GIS architectures	22
3.1	Example of comparison between several unit's scales and resolution units	38
3.2	Example of comparison between angle-based resolution units with dot/meter	42
4.1	Ratio of resolution from one zoom level to the previous one in Google Maps	65
4.2	Resolution on pixel/meter (or dot/meter) for several locations having different latitude values at zoom level equals to 19	66
5.1	Listed of rational cubic spline function's parameters along with their values required in the first experimentation	95
5.2	Control points connecting nrg.crd file and usm.crd file existed in nrg.map file	97
5.3	Control points connecting usm.crd file and penang.crd file stored in usm.map file	97
5.4	Parameters value used in transforming coordinate from usm.crd to penang.crd	98
5.5	Control points connecting penang.crd file and malaysianpeninsular.crd file stored in penang.map file	98
5.6	Parameters value used in transforming coordinate from penang.crd to malaysianpeninsular.crd	98
5.7	Control points connecting malaysianpeninsular.crd file and world/00.crd file stored in malaysianpeninsular.map file	99
5.8	Parameters value used in transforming coordinate from malaysianpeninsular.crd to world/00.crd	99
5.9	Control points connecting nrg-a.crd file and usm-a.crd file stored in nrg-a.map file	100
5.10	Control points connecting usm-a.crd file and penang-a.crd file stored in usm-a.map file	100
5.11	Parameters value used in transforming coordinate from usm-a.crd to penang-a.crd	101
5.12	Control points connecting penang-a.crd file and world/00.crd file stored in penang-a.map file	101

5.13	Parameters value used in transforming coordinate from penang-a.crd to world/00.crd	101
5.14	Control points connecting nrg-b.crd file and usm-b.crd file stored in nrg-b.map file	102
5.15	Control points connecting usm-b.crd file and world/00.crd file stored in usm-b.map file	103
5.16	Parameters value used in transforming coordinate from usm-b.crd to world/00.crd	103
5.17	Several resolution values in GLOBALBASE with their appropriate zoom levels in Google Maps at latitude 5.401862	107
5.18	Control points connecting kokudo20000/coord/bessel.crd file and kokudo20000/coord/00.crd file stored in kokudo20000/coord/bessel-00.map file	113
5.19	Parameters value used in transforming coordinate from kokudo20000/coord/bessel.crd to kokudo20000/coord/00.crd	113
5.20	Control points connecting kokudo20000/coord/00.crd file and world/00.crd file stored in kokudo20000/coord/world.map file	114
5.21	Parameters value used in transforming coordinate from kokudo20000/coord/00.crd to world/00.crd	114

LIST OF FIGURES

	Page
2.1 Raster data model	9
2.2 Vector data model	11
2.3 Client-server architectures. (a) Two-tier client-server architecture. (b) N-tier client-server architecture.	18
2.4 An example of internet GIS architecture	19
2.5 An example of data clearinghouse in an internet GIS	20
2.6 An example of mobile GIS architecture	21
2.7 Distributed architecture in GLOBALBASE	23
2.8 Coordinate system and mapping concept in GLOBALBASE	25
2.9 GLOBALBASE protocol	26
3.1 Two possible cases of two pairs reference points in mapping resources. (a) Case of top-left and down-right reference points. (b) Case of down-left and top-right reference points.	33
3.2 Mapping process using three pairs of reference points	35
3.3 Geometric structure of a globe. (a) Sphere model. (b) Spheroid model.	39
3.4 Difference between geodetic latitude (ϕ) and geocentric latitude (ϕ')	41
3.5 Architecture of the proposed high-level middleware	47
3.6 Overall design of the extended Cosmos	53
4.1 Cosmos shows some overlapped maps	61
4.2 Alteration on X-Y length ratio of a building represented in two different maps	62
4.3 Google Maps showing Penang Island at zoom level equals to 10 in Map type view	64
4.4 Process to transform coordinate of a location from arbitrary coordinate system into latitude-longitude coordinate system in GLOBALBASE	68
4.5 LoD conversion from GLOBALBASE into Google Maps	71
4.6 LoD conversion from Google Maps into GLOBALBASE	74

4.7	Cosmos screenshot in CGB Mode	75
4.8	Cosmos screenshot in CGM Mode	77
5.1	Diagram of installation setup for experimentation environment	83
5.2	Mapping route of spatial data in USM Landscape Server	86
5.3	Two mapping routes of spatial data stored in USM Landscape Server for comparison purpose	88
5.4	Relations between several spatial data existed in Nichibunken Landscape Server	89
5.5	Transformation process on spatial attributes of School of Computer Science from GLOBALBASE into Google Maps	92
5.6	Plot of Google Maps resolution values interpolation (at zoom level equals to 19) using rational cubic spline function	94
5.7	Result of the 1 st experimentation: Cosmos and Firefox browser showing maps having equivalent coordinate points for the first case of the first experiment	104
5.8	Result of the 1 st experimentation: Cosmos and Firefox browser showing maps having equivalent coordinate points for the second case of the first experiment	104
5.9	Result of the 1 st experimentation: Cosmos and Firefox browser showing maps having equivalent coordinate points for the third case of the first experiment	105
5.10	Inaccuracy on the first case of the first experimentation	106
5.11	Result of the 2 nd experimentation: screenshots of Cosmos and Firefox browser showing experimentation results of LoD transformation	108
5.12	Result of the 3 rd experimentation: Cosmos and Firefox browser showing Asian-Pacific region	110
5.13	Result of the 3 rd experimentation: Cosmos and Firefox browser showing Japan	111
5.14	Result of the 3 rd experimentation: Firefox browser displaying spatial data from Google Maps acquired based on spatial attributes defined in kokudo20000/coord/bessel.crd	112
5.15	Result of the 3 rd experimentation: Firefox browser displaying spatial data from Google Maps acquired based on spatial attributes defined in world/00.crd	114

LIST OF ABBREVIATION

ACRP	Auto-Configured Routing Protocol
AJAX	Asynchronous JavaScript XML
API	Application Programming Interface
CGB	Cosmos-GLOBALBASE
CGM	Cosmos-Google Maps
DBMS	Database Management System
DINE	Dual Independent Map Encoding
DOI	Digital Object Identifier
ECEF	Earth Centered Earth Fixed
FGDC	Federal Geographic Data Committee
FMP	Feature Metadata Database
GB	GLOBALBASE
GBIR	GLOBALBASE Input Receiver
GBRG	GLOBALBASE Resolution Generator
GIS	Geographic Information System
GMIR	Google Maps Input Receiver
GSD	Ground Sample Distance
GUI	Graphical User Interface
IP	Internet Protocol
LAN	Local Area Networks
LoD	Level of Detail
NISO	National Information Standards Organization
PMD	Plate Metadata Database
PSE	Partial Search Engine
PURL	Persistent Uniform Resource Locator

R-ZL	Resolution to Zoom Level
RCSI	Rational Cubic Spline Interpolation
TCP	Transmission Control Protocol
URL	Uniform Resource Locator
WAN	Wide Area Networks
WAP	Wireless Application Protocol
XL	XML Lisp
XML	Extensible Markup Language
ZL-R	Zoom Level to Resolution

PENGINTEGRASIAN GIS BERASAS WEB KONVENSIONAL DAN GLOBALBASE MENGGUNAKAN TRANSFORMASI GEOMETRI DAN PERISIAN PERANTARAAN PERINGKAT TINGGI

ABSTRAK

GLOBALBASE, satu aplikasi GIS yang mempunyai seni bina teragih berotomasi, menyediakan satu kaedah baru dalam perkongsian data ruang. Sistem ini tidak mempunyai sekatan yang dapat membawa kepada pembatasan bentuk data ruang. Melalui sistem ini, pembekal data boleh berkongsi data mereka melalui banyak pelayan teragih tanpa sebarang sekatan daripada satu pusat kawalan maklumat ruang. Bagaimanapun, jumlah data ruang yang tersedia terus berkembang dalam satu perkembangan yang pesat. Saat ini, kehadiran aplikasi-aplikasi GIS berasaskan laman web yang dapat dibenamkan membawa satu cara baru dalam memanfaatkan data ruang.

Tesis ini bertujuan untuk mereka bentuk aplikasi GIS yang dapat memperoleh data ruang daripada GIS yang berbeza pentas. Di sini, kita cuba mengukuhkan fungsi-fungsi GLOBALBASE supaya system tersebut boleh mendapatkan dan menyepadukan data ruangnya dengan Google Maps, satu GIS berasaskan web. Untuk mencapai tahapan ini, kesesuaian antara sifat-sifat keruangan pada kedua sistem menjadi suatu keperluan. Tambahan pula, satu kaedah yang membolehkan pelanggan GLOBALBASE untuk mendapatkan dan memaparkan data ruang daripada Google Maps diperlukan.

Usaha kami adalah untuk mencadangkan kaedah-kaedah perubahan pada sifat-sifat keruangan agar kesesuaian antara data ruang dapat diperolehi. Kami turut melakar satu perisian perantaraan peringkat tinggi yang akan membolehkan pelanggan GLOBALBASE, yang disebut sebagai Cosmos, menggunakan data ruang Google Maps. Tambahan pula, kami juga meningkatkan beberapa fungsi tambahan pada Cosmos yang akan membolehkan pelanggan tersebut menerima dan mewakili data ruang Google Maps.

Kami menjalankan beberapa percubaan untuk menilai kaedah yang dicadangkan bersamaan dengan menguji kebolehan perisian perantaraan peringkat tinggi. Hasil-hasil percubaan yang kami perolehi menunjukkan bahawa, secara umum, kaedah-kaedah perubahan tersebut boleh mencapai satu tahap ketepatan tertentu.

INTEGRATION OF CONVENTIONAL WEB-BASED GIS AND GLOBALBASE USING GEOMETRIC TRANSFORMATIONS AND HIGH-LEVEL MIDDLEWARE

ABSTRACT

GLOBALBASE, a GIS application having autonomous distributed as its architecture, provides a new method in sharing spatial data. This system does not have restriction that leads into limitation on spatial format. Through the system, data provider can share their data from numerous distributed servers without any restriction from a spatial information control center. Nevertheless, the number of spatial data existed grows in a rapid acceleration. Currently, the existence embeddable web-based GIS applications bring a new way in benefiting spatial data.

The intent of this thesis is to design GIS application that is able to obtain spatial data from different GIS platforms. Here, we try to enhance GLOBALBASE functions so that system can receive and integrate its spatial data with Google Maps, a web-based GIS. To achieve this state, interoperability between spatial attributes on both systems is become a requirement. Moreover, a method enabling GLOBALBASE client to obtain and represent spatial data from Google Maps is necessitated.

Our attempt is to propose transformation methods on spatial attributes so that interoperability between spatial data can be obtained. We also design a high-level middleware that will enable GLOBALBASE client called Cosmos accessing the Google Maps spatial data. In addition, we also enhance Cosmos with several additional functions that will enable the client receiving and representing Google Maps spatial data.

We conduct several experiments to evaluate the proposed methods along with to test the ability of the high-level middleware. The experimental results we obtained show that, in general, the transformation methods can meet a certain level of accuracy.

CHAPTER 1

INTRODUCTION

1.1 Background

Generally, phenomena happened in the real world can be described, represented, and modeled by using two kinds of data which are spatial data and non-spatial data. The spatial data is used to represent the spatial aspects of the phenomena. These kinds of data can solve queries such as to determine distance between two locations or to find the shortest path from one place to another. Meanwhile, the purpose of non-spatial data, also known as attribute data, is to represent descriptive aspects such as items or properties of the modeled phenomena up to its time dimension of an entity.

There are several systems that are built to manage data. Database management system (DBMS) is one of the systems which often be found in our daily life. Although this system has already become an inseparable role in many sectors of human activities, yet this system is only able in representing and solving non-spatial queries. Because of that, the existence of system that capable in processing both spatial and non-spatial data is essential. Geographic information system (GIS) was built in order to fulfill this necessity. To simplify, GIS is an information system to store, manipulate, analyze, and represent any real world's phenomena with both spatial and non-spatial data.

As time goes by, expectation toward GIS technology is increased. Its ability in spatial data management makes GIS possible to integrate any kinds of information into a common spatial and visual language. This is very important in order to solve problems that encompass some fields since nowadays society is evolving toward greater specialization where people have become accustomed to see the world in pieces depend on their own discipline and specialization.

In the meantime, real world is changing at an ever increasing rate each second. The changing might be caused by human activities along with its side effects such as city development, alteration of land use from forest into commercial agricultural enterprise, or global climate change; by natural cause (since the earth's surface is not static); or by disasters happened like earthquake, volcano eruptions, and landslide. These alterations in the real world landscape imply that geographical data change acceleratingly. Therefore, satellite as one of the tools in obtaining and providing data can produce terabytes of spatial data per day, in addition to accuracy level of the produced data. At the same time, an up to date data will always be required to get an accurate and reliable information system. This is one of the reasons why the amount of spatial data existed is really huge. Moreover, these large volumes of data are provided by many parties scattered all around the world.

The above situations lead into problems for the existing GIS architecture in order to represent a complete spatial data and present a challenge for the development of computer system in order to store, analyze, manipulate, and represent these huge data sets. Nowadays, most of GIS applications are implementing conventional centralized architecture. This architecture stores its data in one server or in some servers with a central server as its spatial information control center that will handle all of the searching and overlapping process through a clearinghouse method. These systems only permit certain parties to import and modify the spatial data which give rise to restriction in spatial data sharing. Such thing become obstacles in GIS growth since centralized architecture also leads the system into a bottleneck problems and data sharing is an important factor in order to provide the user with up to date and complete geographical information.

GLOBALBASE (Mori, 2004) is a GIS application that implements autonomous distributed architecture, connecting spatial information that is delivered by data provider from numerous servers. In GLOBALBASE architecture, data providers can build their own

server where later on will be connected to the other providers' servers. As a decentralized system, these servers will work without any restriction from a spatial information control center. By starting up their own server, data providers can upload their own spatial data freely so it is possible to share individual information through the GIS application. Users can also access and retrieve spatial data from numerous GLOBALBASE server existed in the network directly. Thus, the architecture avoids the bottlenecks problems while enables spatial data sharing from numerous parties. Yet, in order for GLOBALBASE system able to share this individual information, they employ exclusive data structure and protocol in addition to the autonomous distributed architecture.

1.2 Problem Statement

At this moment, GLOBALBASE is able to get and share geographical information from its own platform. This means that GLOBALBASE is only able to manage and integrate spatial data supplied under GLOBALBASE community only. Meanwhile, many other geographical information are made and supplied using other GIS applications. Some organizations focus in using GIS application only to make spatial data in a specific field. Therefore, interoperability between GIS applications is become important in order to collect as much as geographical information in less amount of time so that users can make an integrated and balanced solution from all points of view of their problems.

1.3 Research Objectives

This research work intends to design GIS application that is able to obtain spatial data from different GIS platforms. To be more specific, we try to enable GLOBALBASE in obtaining spatial data from another GIS application. Here, we will integrate GLOBALBASE with a web-based GIS called Google Maps

In order to achieve our goal, there are two main problems need to be resolved. One is the variety of spatial data formats used by GIS applications, although some GIS applications

employ similar standard. In our case, both GLOBALBASE and Google Maps have their own standard on their spatial data format. Thus, to attain interoperability of spatial data, we develop methods to transform spatial data format from one into another. The transformation methods will be applicable into any kinds of spatial data format.

Another problem is the difference of architectures applied on GLOBALBASE and Google Maps. The diversity of architecture leads into different method in order for these GIS applications to obtain their spatial data, although GLOBALBASE and Google Maps work on top of standard internet protocol. To solve this matter, we design a high-level middleware connecting the autonomous distributed GIS and web-based GIS applications, including Google Maps.

1.3 Contribution of This Thesis

This thesis contributes towards the following:

- Methods on transforming spatial attributes from one GIS application format into another. Generally, the transformation methods can be divided into coordinate transformation and level of detail conversion. Through these transformations, we hope spatial attributes interoperability between GIS applications can be achieved.
- Proposing a high-level middleware to enable GIS applications from different platforms and architectures share and communicate one another. As mentioned in the previous section, we choose a distributed GIS called GLOBALBASE and a web-based GIS called Google Maps as our case study. Through these middleware, spatial data can be shared without any necessity in transferring the spatial data itself.
- Proposing a design of enhanced GLOBALBASE client. The design will enable the new GLOBALBASE client receives and represents spatial data from Google Maps.

1.4 Thesis Outline

This thesis is organized into six chapters. The contents are arranged such that each previous chapter provides a basic idea to further proceed to the next chapter. This first chapter introduces the background of our work along with our research objectives and contributions.

The second chapter explains literature review and fundamental concepts related to our work and issues surrounding it. Some basic GIS definitions and concepts, including GIS data model, GIS architectures, and some fundamental geodesy concepts, are described. We also introduce fundamental concepts of GLOBALBASE system, as it will be our study case in this work.

Chapter 3 becomes the main chapter in this thesis. This chapter contains our contributions on this research. In the first section, we will represent several transformation methods that can be utilized in order to acquire spatial data interoperability from one format into another. The second section will describe the proposed high-level middleware intermediating GLOBALBASE and Google Maps. Meanwhile, the new GLOBALBASE client design will be proposed in the third section.

Chapter 4 contains of implementation details of the work. Here, in the first section of this chapter, we describe implementation of proposed transformation methods in obtaining spatial data interoperability between GLOBALBASE and Google Maps. We also introduce the new graphical user interface for the GLOBALBASE client, as the new client need to be able to acquire and represent spatial data from Google Maps, in the second section.

Chapter 5 will be divided into two parts. The first part will describe about the experimentations we carry out in order to examine the proposed transformation methods

along with the high-level middleware explained in Chapter 3. Result of the experimentations will be represented in the second part of this chapter.

In Chapter 6, we represent summary of each chapter in this thesis. We also revisit our research contributions with regard to methods we proposed in Chapter 3 and its results in Chapter 5. In last part of this chapter, we present a discussion and suggestion for future work related to this research.

CHAPTER 2

LITERATURE REVIEW

This chapter mainly presents literature review on fundamental concepts of GIS along with a brief description on GLOBALBASE, as our object in this work. In Section 2.1, we present definitions of GIS from several points of view. Section 2.2 describes spatial data models in GIS while Section 2.3 presents fundamental concepts in geodesy, which is a scientific discipline concerning the earth surface and mapping concepts used in GIS. In Section 2.4, several architectures of GIS are described. Finally, basic concept in GLOBALBASE system is presented in Section 2.5 before a summary of this chapter closes this chapter.

2.1 Definition of Geographic Information System

It is difficult to define the term geographic information system, usually abbreviated by GIS, since it represents the integration of numerous areas just as the field of geography. Yet by referring to the main elements of GIS, which are “geographic”, “information”, and “system”, a GIS can be said as a particular form of information system applied to geographical data (Waits). It is a special case of information system where the database consists of observations on spatially distributed features, activities, or events, which are definable in space as points, lines, or areas. Geographic information system manipulates data about these points, lines, and areas to retrieve data for ad hoc queries and analyses (Tiglaou, 2005).

According to National Centre of Geographic Information and Analysis, geographic information system is a system of hardware, software and procedures to facilitate the management, manipulation, analysis, modeling, representation and display of georeferenced

data to solve complex problems regarding planning and management of resources (NCGIA, 1990) (Escobar *et al.*, 2001).

Some other definitions from different literatures are a computer system, which is used for capturing, storing, checking, integrating, manipulating, analyzing, and viewing data that related with positions in the earth surface (Prahasta, 2001) or a computer-based system for storing, maintaining, querying and analyzing geographic data (Hammerie, 1996).

Up till now, most of GIS definitions from all sorts of literatures are still general, incomplete and elastic (Prahasta, 2001). GIS definition will always expand, increase, and have more variations, in accordance with geographic information system development. Yet it is clear that the main aspect that distinguishes GIS from another system is their ability to display geographic or cartographic data and to analyze this data in a very flexible way (Hammerie, 1996).

2.2 Spatial Data Model in GIS

The main difference between GIS and the other systems that manage data is its ability in storing, processing, analyzing, and displaying spatial data. Spatial data or sometimes called as geographic data occupy a position on the earth's surface (Hammerie, 1996). In traditional GIS that deal with two-dimensional spatial data, the type of spatial objects can be divided into three categories; point data, linear data, and polygon data. In managing these data, there are two kinds of spatial data model exist, depending on the type of features they represent and the purpose to which the data will be applied, which are raster data model and vector data model. In addition to documenting these data models, existed spatial-based information as properties of raster and vector data called metadata.

2.2.1 Raster Data Model

Raster data model is a method to storage, process and display spatial data using matrixes structure (Unimelb) (Escobar *et al.*). The datasets are composed of rectangular arrays of regularly spaced square grid cells, obtained by dividing the area into rows and columns. Each cell has a value that representing a property or attribute of interest. The accuracy of the data really depends on the resolutions or the size of its pixel compared to the real-world surface, the acquisition device and the quality of the original data source (Prahasta, 2001) (Wu, 1999). These kinds of datasets are especially suited to the representation of continuous data, such as elevation, soil pH, and salinity in water (Hurvitz, 2004). Figure 2.1 shows a diagrammatic model of raster datasets in representing real-world features.

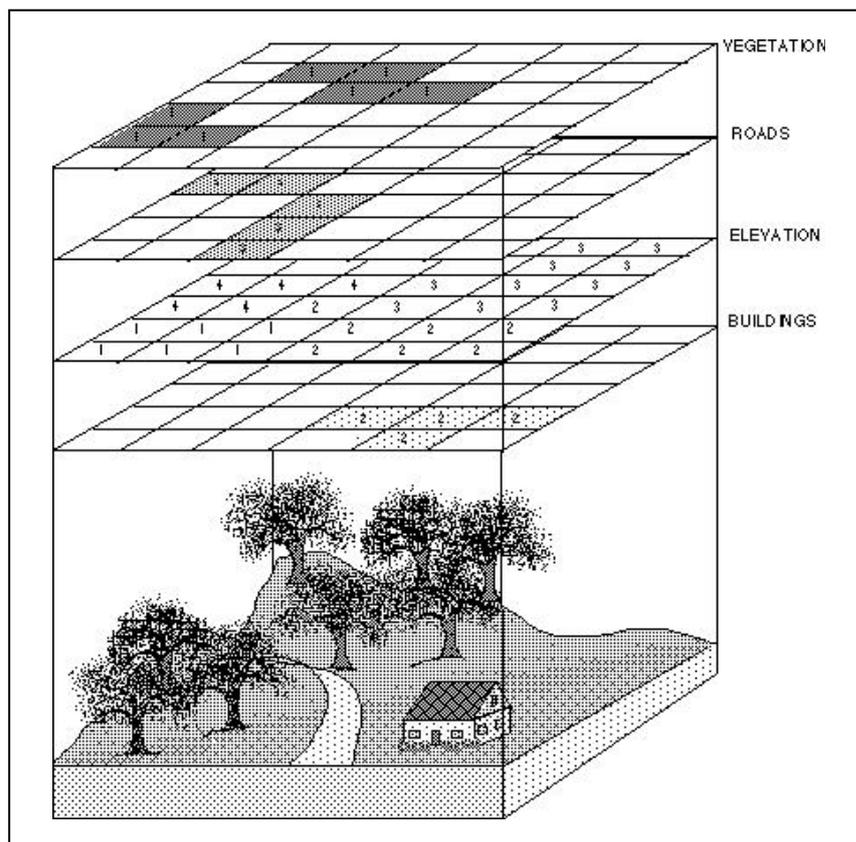


Figure 2.1: Raster data model (Hurvitz, 2004)

Raster datasets are spatially referenced by using one of its corners of the raster layer (usually the upper-left corner or the lower-left corner) as a georeference. Since cell size is

constant in both X and Y directions, cell locations are referenced by row/column designations. Each cell (a user-defined area representing phenomenon) or pixel (the smallest resolvable piece of a scanned image) contains a value representing some numerical phenomenon or a code use for referencing to a non-numerical value, such as color, elevation, or an ID number. There are three methods in representing the value of a pixel and it is produced with different sampling process, which are 1) by using the mean value of the area represented; 2) using the sampling value located in the centre of the pixel; 3) using the sample value located in the corners of the grid. Each sampling process has its own advantages and disadvantages. In the implementation, there are four types of raster data architecture, which are chain coding, run-length encoding, block coding, and quad tree where each type have different level of data configuration and package to encourage efficiency (Oluseyi, 2002).

Compared to vector data model, raster data structure is much simpler and easier to manipulate with simple mathematical functions. This data model is compatible with remote sensing satellite images and any scanned spatial data images. It is also easier to overlaying and combining raster data spatial with remote sensing data. Besides that, this kind of spatial data is usually more up to date than the vector data in the same area since the technology, methodology, and procedure to get and manipulate raster image is easier and cheaper (Prahasta, 2001).

2.2.2 Vector Data Model

Vector data model is another method in managing spatial data that represent features as discrete points, lines (or curves), and polygons that are geometrically and mathematically associated. One of the strengths of the vector data model is that it can be used to render geographic features with great precision. However, this comes at the cost of greater complexity in data structures, which sometimes translates to slow processing speed. Figure 2.2 shows a diagrammatic model of vector datasets.

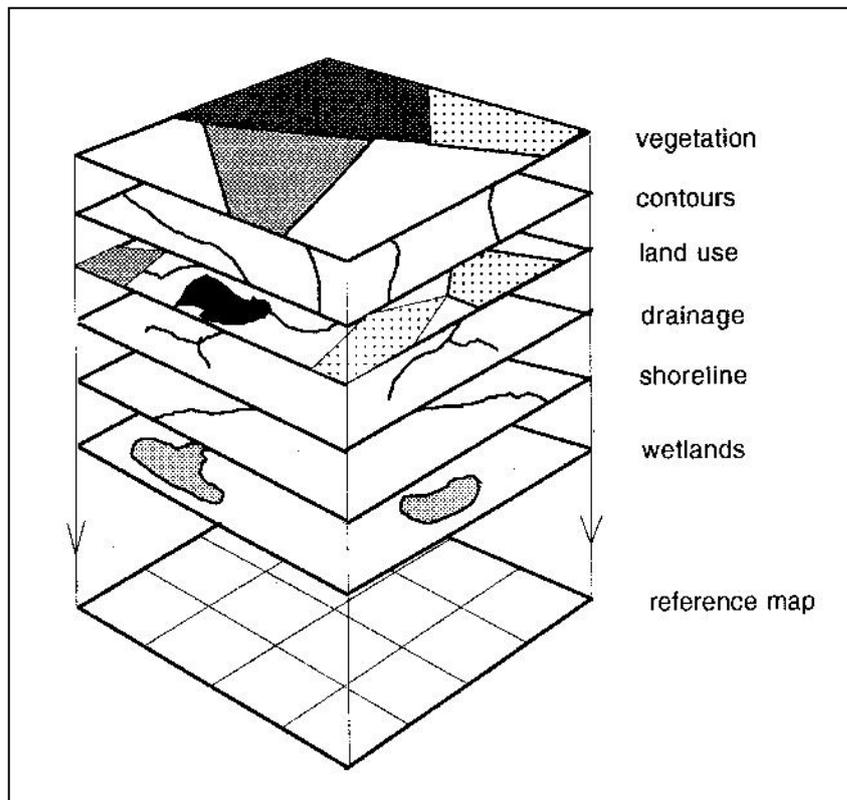


Figure 2.2: Vector data model (Meaden and Chi, 1996)

The spatial data is defined in two dimensional Cartesian coordinate system (x and y). Points are represented with a pair of coordinates (x, y) and lines (or curves) are stored as a series of point connected each other, where each pair represents a straight-line segment (Wu, 1999). Meanwhile, polygons are stored as a list of points where the polygon's ending point has the same coordinate with its starting point. Vector data environment are divided into different types of data architecture such as whole polygon structure, Dual Independent Map Encoding (DIME) file structure, arc-node structure, relational structure, and digital line graph (Oluseyi, 2002).

Vector data model will need less disk storage compared to raster data model since raster images need space for all pixels while only point coordinates are stored in vector representation (Wu, 1999). Numerous thematic maps can be derived by combining various

attributes in a layer. Using this data model, topology and network connection, coordinate transformation and projection can be done easily and it also has a high spatial resolution. Moreover, since the boundary in vector data model is clearly visible, it is very suitable to make administration and parcel of privately owned land maps (Prahasta, 2001).

2.2.3 Metadata in GIS

Term of metadata is used differently in various communities. In brief, metadata are often defined as data about data or information that gives description about data. From this definition, metadata are also data, depends on the point of view. Therefore, it might possible to make metadata about metadata (meta-metadata), metadata about metadata about metadata (meta-meta-metadata), and so on. Metadata give documentation about data set (resource) so that the data can be managed, maintained, and utilized properly either by the data provider or the user.

Providing documentation by equipping metadata into data has become important issue, especially in data management. Metadata are able to facilitate, speed up and enrich relevant resource discovery by using proper criteria, organize electronic resources, bridge semantic gap to provide interoperability and legacy resource integration. Metadata also provide digital identification using unique standard number, file name, Uniform Resource Locator (URL), or persistent identifier such as Persistent URL (PURL) or Digital Object Identifier (DOI). Other metadata purposes are to support resource archiving and preservation, to optimize compression algorithm, to enable variable content presentation and to automate workflows.

According to NISO (National Information Standards Organization), there are three main type of metadata. They are descriptive metadata, structural metadata, and administrative metadata. Descriptive metadata give information about a resource so that the resource is easier to be discovered and identified, including its title, author, keywords, and

abstract. Structural metadata are used in documenting composition of compounded objects. Meanwhile, administrative metadata provide information that are used in resource management, such as file type, creation time, and procedure, accessibility, and other technical information (NISO, 2004).

Concerning its storage, metadata either can be stored in the same file as the data (internally) or in a separate file (externally), depending on the requirements since both have their own advantages and disadvantages. Storing metadata internally allows data transfer with its metadata at once. This can ensure the metadata will not be lost, obviate linking problems with the data, and make the metadata easier to be manipulated and updated along with the data. Yet, internal storage creates high redundancy and does not allow holding metadata together. Meanwhile, external storage allows bundling metadata (usually in a database) so that it will be more lenient in managing, more efficient in searching, no redundancy, and metadata can be transferred simultaneously when using streaming. However, the way metadata are linked to the data through this method becomes an important matter to be considered.

In term of data format, storing metadata in a human-readable format such as XML can be useful because users can understand and edit them without any tools at all. On the other side, these formats are not optimized for storage capacity, i.e. it may be useful to store them in a binary non-human-readable format instead to speed up transfer and save memory.

In GIS field, metadata is used as resource in data discovery, data transfer, data management, and data use. It is important to equip spatial data with metadata in standard format so that the discovering can be done in uniform on each application and to ascertain reliability on spatial data between GIS applications in the sharing process. For this, Federal (U.S.) Geographic Data Committee (FGDC) has developed standard for metadata used to describe a data set which are identification information, data quality information, spatial data

organization information, spatial reference information, entity and attribute information, distribution information, metadata reference information, citation information, time period information, and contact information (Harmon & Anderson, 2003). Nevertheless, there might be priorities between these elements to be stored by differentiate information as mandatory or optional (Bernhardsen, 2002).

2.3 Geodesy Concepts

Geodesy is an inseparable subject with GIS. It is a scientific discipline that study about the shape and size of the earth, determining point location, length and direction of lines measurement (geometric geodesy), in addition to examining earth's gravitational field (physical geodesy). This subject is important in describing the shape of earth surface, including the geographical position of its elements. Some basic concepts that are used in geodesy are geoid, reference ellipsoid, geodetic datum, map projection, and coordinate system.

The earth physical surface is highly irregular and constantly changes. Therefore, a computational surface is needed to representing the real-world surface and also for geodetic calculation in determining point's location, distance and direction in the real-world surface. Reference ellipsoid is used for this purpose. It has the same size (volume) as the geoid (an essential figure of the earth, an equipotential surface which approximately coincides with the mean ocean surface) and defined by its equatorial radius and its flattening. However, the ellipsoid's surface can be considered as a spherical surface when analyzing an area less than 100km² and as a plane surface when working with an area less than 55km² (Prahasta, 2001) (Wikipedia).

Geodetic datum is a set of constant values used to define coordinate system for geodetic control. Nowadays, to define geodetic datum, at least eight constant values are needed; three constant values to define coordinate system original point (X_0 , Y_0 , Z_0), three

constant values to determine coordinate system direction, and two other constant values to define ellipsoid's dimension that is used (its equatorial radius and its flattening values) (Prahasta, 2001). Hundreds of different datums have been used to frame position descriptions (Dana, 1995). It can range from flat-earth models used for plane surveying to complex models used for international applications which completely describe the size, shape, orientation, gravity field, and angular velocity of the earth. According to its reference ellipsoid, geodetic datum can be divided into local datum, regional datum, and global datum. Geodetic datum also can be divided into horizontal datum and vertical datum (Prahasta, 2001).

A map projection translates the locations on the globe onto the flat surface of the map. It transforms coordinates from geographic coordinates (latitude and longitude of the earth) to projection coordinates (x and y in GIS) (Teng *et al.*, 2004). All map projections distort the shapes of the features being displayed to some degree, as well as measurements of area, distance, and direction. Therefore, an appropriate technique of map projection shall be used in minimizing these distortions.

A coordinate system is collection of rules that determine how relevant coordinates represent points. The rules usually define the original point along with some coordinate axes used to measure distance and angle to produce coordinates. Coordinate system could be grouped according to how the original point is located (geocentric, topocentric, heliocentric, etc.), type of surface that is used as reference model (flat, sphere, and ellipsoid), and its axes direction (horizontal and vertical) (Prahasta, 2001).

2.4 Architecture of GIS

GIS development cannot be separated with computer technology. This information system has been implemented through various technology since computer was applied as analytical and display tools in geographic area for the first time (Clarke, 1996), depends on

contemporary computer architectures (ESRI, 2003). Thus, GIS architectures had evolved from traditional and centralized architecture, such as mainframe GIS, to desktop GIS and currently to distributed architectures accustomed by most of current GIS applications. All of these architectures are differentiated by how the systems are executed.

Basically, GIS has three main subsystems, which are data management (data storage), spatial operations and analysis (data processing), and mapping (data presentation). Data management is a subsystem used to read and write spatial and non-spatial data into permanent storage media. The spatial operations and analysis implements a series of geographic analysis functions such as proximity and overlay analysis, data conversion, grid analysis, and three-dimensional analysis. Mapping subsystem deals with geographic information manipulation and visualization on the graphical user interface (GUI) and display area where user interacts with the system. This subsystem includes functions such as projection and datum transformation, map displays, and two-dimensional and three dimensional visualization tools (ESRI, 2003) (GIS Lounge, 2000).

All of these subsystems above can be executed in a single machine (single-tier architecture) or in a client-server architecture having two or more machines. Mainframe GIS architecture adopts monolithic computing model where all of the subsystems above are executed in the same mainframe computers and accessed by using dumb terminals over Local-Area Networks (LANs). Meanwhile, in desktop GIS, the GIS programs are placed in desktop computers. Both of these architectures can be categorized as centralized GIS. This kind of architecture leads into several problems hampering GIS development and spatial data dissemination.

GIS application having client-server architecture is called distributed GIS. According to Sommerville (2001), a distributed system is a system where the information processing is distributed over several computers rather than confined to a single machine.

There are various distributed GIS architectures existed. Each of the architecture is differentiated by the number of tier used on the system and how the distributed components communicate each other.

2.4.1 Client-Server Architecture

The simplest client-server architecture is two-tier architecture which popular in early 1990s. This architecture disaggregated GIS functions into server side and client side. Therefore, this kind of architecture can be classified into two forms; thin-client model and thick-client model. Thin-client model relies on server side to handle processing and data management. Client is simply responsible in data presentation. In contrary with thin-client model, thick-client model implement logic processing and data presentation on the client side while server only needs to manage data.

Each model has its own advantages and disadvantages. Thin-client is simple but will place heavy task on server and network leads into scalability and performance problems. Meanwhile, thick-client distributes processing more effectively but its system management is more complex (Sommerville, 2001). Yet, two-tier architectures have their own disadvantage since these models only provides clear distinction between data management and user interface responsibilities but fail to define place for logical operations and analysis clearly (GIS Lounge, 2000). Therefore, this architecture brings problems in upgrading and extending the system (ESRI, 2003).

Most recent, three-tier architecture has been utilized as an alternative approach in client-server architecture (GIS Lounge, 2000). Here, three main functions of the system are done in separate processes. Client side implements user interface, server side has responsibility for data management, and middle part applies the logical operations and analysis component. The middle part can be disaggregated into several parts depends on the number of the logical process. This kind of architecture is called as N-tier architecture where

N reflects number of intermediate parts between client and server. With these architectures, systems are more scalable and reduce more network traffic compared with two-tier architecture. Figure 2.3 shows architectures of client-server GIS.

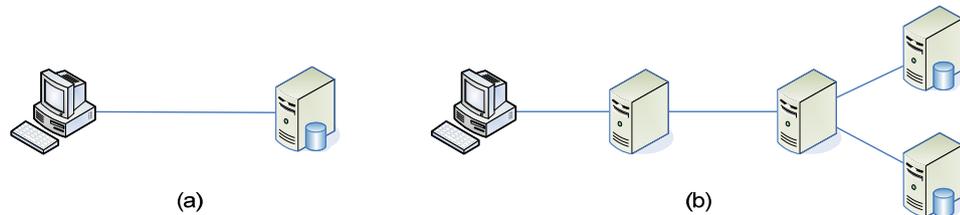


Figure 2.3: Client-server architectures. (a) Two-tier client-server architecture.

(b) N-tier client-server architecture.

2.4.2 Internet GIS

Basically, internet GIS is based on client-server GIS architecture. Nevertheless, its components on the system are connected through an internet. Internet involvement affects behaviors of GIS in three major areas, which are spatial information dissemination, GIS process model, and data access. Internet enables dissemination of spatial information along with GIS analysis results to a wider audience. This is due to several internet GIS use web browser as their client and thus spatial information and the GIS analysis results can be explored by public generally. In addition, by the evolvement of internet and web technologies recently, internet GIS focuses more on applications on the web client side by using GIS plug-ins, ActiveX controls, and Java applets, providing better interactions between users and the GIS on internet (Peng, 1997). Figure 2.4 below shows an example of internet GIS architecture.

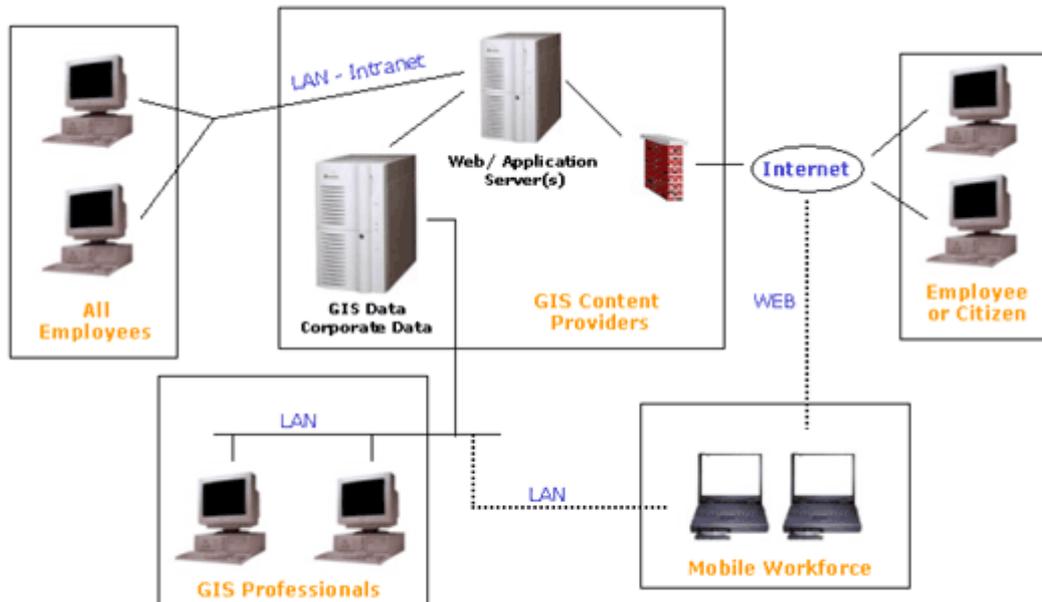


Figure 2.4: An example of internet GIS architecture (Geobusiness)

Internet also enhances the accessibility and reusability of GIS analysis tools by dynamically downloading and uploading GIS processing components. In addition, internet provides user easy access in acquiring GIS data. There are two common systems to access spatial data in internet, which are GIS data clearinghouse and digital libraries.

Clearinghouse or warehouse is a system used to search and access spatial data requested by users without consideration of location in the distributed environment. The searching process is done by comparing users request with stored metadata providing documentation about existing spatial data. Therefore, this system also needs to allow spatial data provider to supply documentation describing their data. Once users found the appropriate spatial data, clearinghouse system will provide network connection for users to access the data (Oh *et al.*). Figure 2.5 shows an example of internet GIS having clearinghouse.

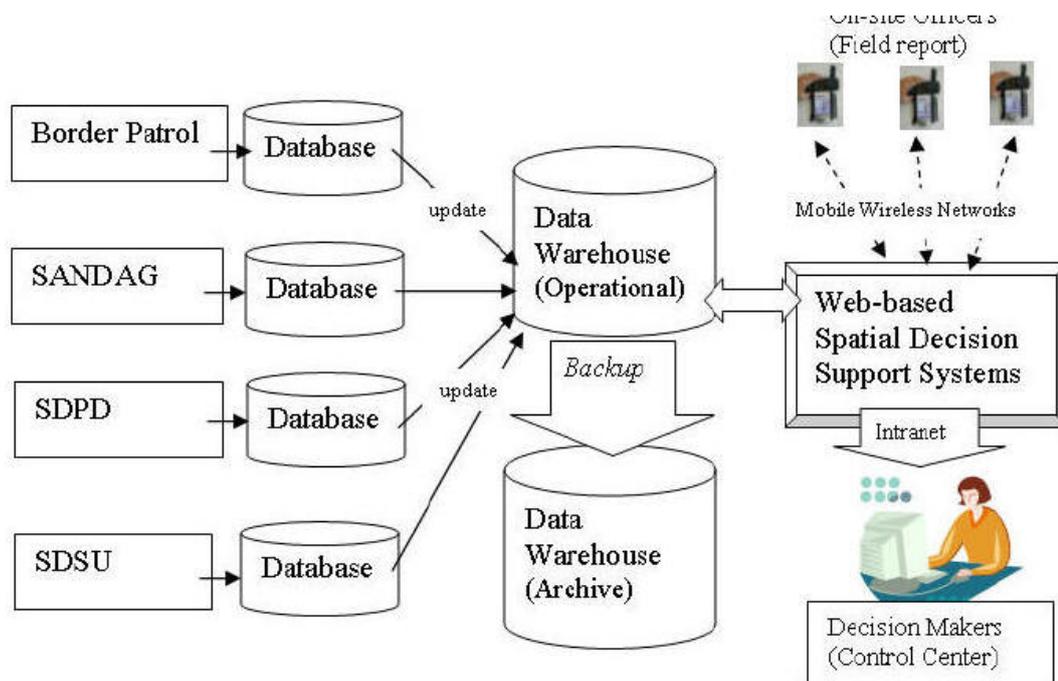


Figure 2.5: An example of data clearinghouse in an internet GIS (SDSU, 2007)

Meanwhile, digital libraries are servers that provides services for querying and downloading of data from the library in addition to services for processing the data before downloading by using high-performance compute servers. This kind of Internet GIS architecture is particularly useful if the amount of data to be processed is very large while the result is relatively small, especially if the data is obtained from a wide-area with relatively low-bandwidth network. Hence, it will be more efficient if the user only has to download the result than download the large input data set and process it locally (Coddington *et al.*, 1999).

2.4.3 Mobile GIS

Mobile GIS is distributed GIS architecture based on mobile computing and mobile internet. This architecture is an extension of internet GIS to mobile internet such as wireless internet/intranet and mobile communication networks. Thus, it can be seen as a sub-division in web-based GIS (Li, 2004). Nonetheless, mobile GIS architectures are different and need adjustment. These are due to limited bandwidth compared to common networks, which leads into several bottlenecks, availability of wireless internet, the diversity of mobile devices with

limitation on their processing power and screen display, and platform diversity of the mobile system (Wang). Figure 2.6 below shows an example of mobile GIS architecture and its components.

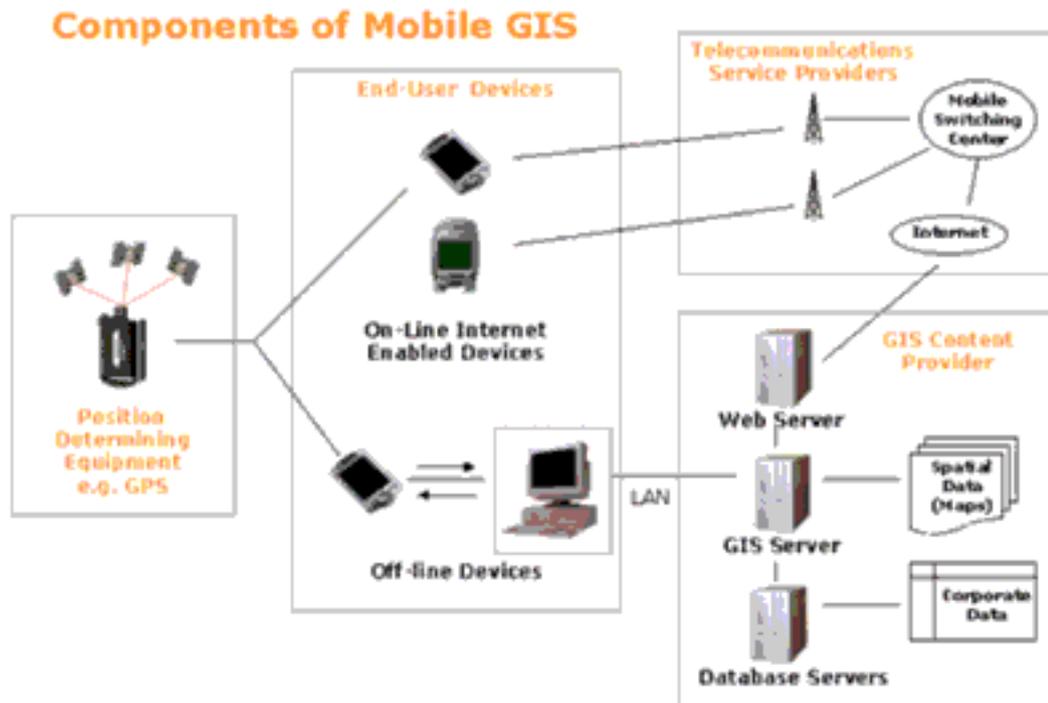


Figure 2.6: An example of mobile GIS architecture (Chen *et al.*).

Between various GIS architectures we have described above, Peng (2003) has tried to compare characteristics of each architecture as shown in Table 2.1 below.

Table 2.1: Characteristics of several GIS architectures

Application Characteristics	Mainframe GIS	Desktop GIS	Distributed GIS	
			Internet GIS	Mobile GIS
Architectural models	Monolithic	Ethernet era client-server (two-tier)	Web-based client-server (three-tier or N-tier)	Wireless client-server (three-tier or N-tier)
Client	Dumb terminal	Desktop computers	Web client	Wireless devices
Client interface	-	Fat Graphic User Interface (GUI) clients	Web browser, Java Beans, ActiveX controls	Mini browser, Wireless Application Protocol (WAP)
Networks	Local Area Networks (LANs)	LANs or Wide Area Networks (WANs)	The internet	Wireless networks and the internet
Server	Mainframe	Application servers and data servers	Web servers, application server, GIS server, and data servers	Gateway server, Web server and GIS servers
Number of accessible server	One	One or a limited few	Thousands or more	Thousands or more

Resources: Internet GIS: Distributed Geographic Information (Peng, 2003).

2.5 GLOBALBASE

GLOBALBASE is a GIS application that uses an autonomous distributed architecture without any central server as its architecture. It is an application that is made in order to solve the qualitative bottleneck happened in any centralized GIS. This system allows data providers to share their spatial information by simply put it up on to a server that can be set up by themselves and also allows users to access multiple servers directly and view the maps stored on the servers. In order to do so, GLOBALBASE employ a special architecture, data structure, and protocol.

2.5.1 GLOBALBASE Architecture

GLOBALBASE has a unique architecture compared to the other GIS architectures. All of the current GIS architectures employ a single coordinate system, leading to only suitable spatial data can be fused into the system. Meanwhile, spatial data available are based on many kinds of coordinate system. Therefore, using these kinds of architecture will limit the number of spatial data that can be put up into the system and this cause qualitative

bottleneck in GIS development. To solve this problem and to decentralize the system, GLOBALBASE use an architecture that does not lean on a single coordinate system. This GIS application can handle any coordinate systems defined individually by data provider. With this, any data provider can put up and share their geographical data so the limitedness problem of geographical information in GIS can be solved. Figure 2.7 shows the architecture that utilized in GLOBALBASE system.

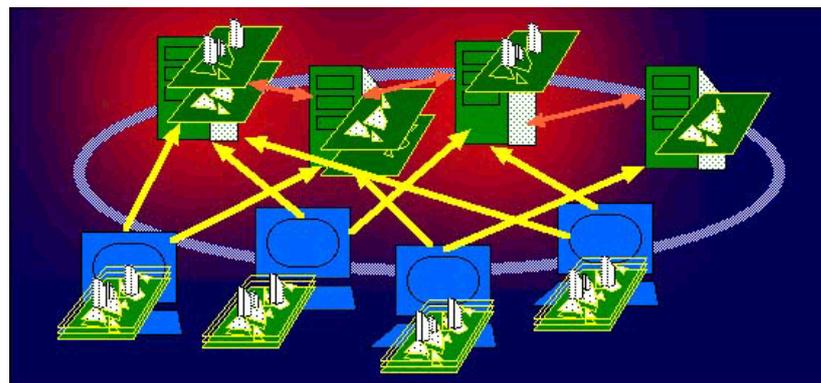


Figure 2.7: Distributed architecture in GLOBALBASE (Mori)

In order to be able to representing the spatial data into one geospace, all of the different coordinate systems defined by the user in GLOBALBASE are integrated using the mapping concept. A mapping plays the role of a bi-directional link between two coordinate systems by connecting some points on a coordinate system into their related pairs in the other coordinate system. Mappings and coordinate systems can be placed in any server since the mappings work as links beyond servers.

The arbitrary coordinate systems defined by data providers can cause maps on the system become cluttered and all of the information on a displayed area will be shown when user zooms on a certain location. It might be difficult to get specific information on the area with this condition. Therefore, to display the information in an orderly manner so that only necessary information is obtained, GLOBALBASE uses distributed search engine called

local search engine. Local search engine utilize coordinate system exist in the spatial data to create an infinite number of small search engines that gather local information.

In the process, browser firstly connects to a local search engine in the area it is currently displaying. When the user enters certain search conditions, the browser searches for geographical information of the surrounding areas. Then, the local search engine will return all geographical information matches with the conditions from a slightly wider area than the displayed area. The browser caches the information and displays it as new information when the user moves to the left or right. If the displayed area changes, the browser will search for a new local search engine corresponding to the new displayed area and search for information based on the user's conditions again. The process will be repeated regularly and the browser will move from one local search engine to another. Moreover, since a server only accumulates local information in this manner, each local search engine can be very simple and does not need to have a large capacity.

In this distributed architecture, two coordinate systems that are superimposed each other cannot be guaranteed if the coordinate systems are connected with one mapping since the coordinate systems are relative and their mapping process are defined arbitrary. Therefore, GLOBALBASE use network routing concept to check the physical relationship between these coordinate systems. Each coordinate system is detected by a unique address assigned automatically using ACRP (Auto-Confabulated Routing Protocol).

2.5.2 GLOBALBASE Data Structure

To fulfill browsing functionality in the distributed architecture discussed above, GLOBALBASE uses a particular data structure. In GLOBALBASE system, data is managed by resources that contain information either created by data provider or added automatically by the server itself called management information. The basic resources in the project consist of coordinate resources, object resources and mapping resources. Each resource is equipped