REVIEW OF AVAILABILITY AND ACCESSIBILITY OF GEOSPATIAL DATA IN THE GREATER MEKONG SUBREGION

A consultancy report in support of the Subregional Environmental Monitoring and Information Systems II Project (SEMIS II)

Prepared by Michael Starbuck
United States Geological Survey

March 2001

TABLE OF CONTENTS

			page
1.0	INTRODUCTION		1
2.0	BACKGROUND AND OBJECTIVES		2
	2.1 Objective of this activity		3
	2.2 Approach		3
3.0	REVIEW OF EARLIER PROJECTS		3
	3.1 Sub-regional Environmental Monitoring and Information System (SEMIS I)		3
		3.1.1 Output 1 – Core dataset definition	4
		Topography and hydrography	5
		Data collection issues	5
		Framework perspective	5
		Additional core data layers	6
		Responsibility for data layers	6
		3.1.2 Output 2 – Spatial database design	7
		3.1.3 Output 3 – Data standard for information to be held in national and sub-regional databases	8
		Spatial data exchange format standard	9
		3.1.4 Output 4 – Metadata standards for information to be held in National and sub-regional databases	9
		3.1.5 Output 5 – Catalogue of existing environmental and natural resources data holding amongst countries of the GMS	9
	3.2 Strategic Environment Framework for the GMS (SEF project)		10
		3.2.1 Early Warning and Information System (EWIS)	11
4.0	REVIEW OF AVAILABLE DIGITAL GEOSPATIAL DATASETS		12
	4.1 Map accuracy issues		12
	4.2 Map projections and coordinate systems		12
	4.3 Recommended data scales/resolutions		13
	4.4 Data accessibility		13
5.0	DATA GAPS		14
6.0	CONCLUSION / RECOMMENDATIONS		14

TABLES

page
Table 1 – Datasets used by EWIS 11

APPENDICES

APPENDIX – 1: Thailand Available Datasets 16

APPENDIX – 2: Vietnam Available Datasets 17

APPENDIX – 3: Cambodia Available Datasets 19

APPENDIX – 4: Lao Available Datasets 21

APPENDIX – 5: United States National Map Accuracy Standards 23

Abbreviations

ADB	Asian Development Bank
ASCII	American Standard Code for Information Interchange
CD-ROM	Compact Disc – Read Only Memory
CIESIN	Center for International Earth Science Information
DEQP	Department of Environmental Quality Promotion
ESRI	Environmental Systems Research Institute
EWIS	Early Warning Information System
FAO	Food and Agriculture Organization
FGDC	Federal Geographic Data Commission
GIS	Geographic Information System
GMS	Greater Mekong Subregion
GRID	Global Resource Information Database
ICIMOD	International Center for Integrated Mountain Development
IIMI	International Irrigation Management Institute
ISO	International Standards Organization
LIDAR	Light Detection and Ranging
Landsat ETM	Earth Resources Satellite (Enhanced Thematic Mapper)
MRC	Mekong River Commission
NAMRIA	National Mapping and Research Information Authority
NASA	National Aeronautics and Space Administration
NEA	National Environment Agency
RRC.AP	Regional Resource Center for Asia and the Pacific
SDTS	Spatial Data Transfer Standard
SEF	Strategic Environmental Framework
SEMIS	Subregional Environmental Monitoring and Information System
SPREP	South Pacific Regional Environmental Programme
TA	Technical Assistance
TIFF	Tagged Image File Format
UNEP	United Nations Environment Programme
USGS	United States Geological Survey
WWF	World Wide Fund for Nature

1.0 Introduction

The Asian Development Bank (ADB) is implementing a Technical Assistance (TA) for the Subregional Environmental Monitoring and Information Systems – Phase II (SEMIS II) in collaboration with the United Nations Environment Programme – Regional Resource Center for Asia and the Pacific (referred to as the Center). The project is co-financed by ADB (through the Japan Special Fund and the Government of Norway) and the Center.

A follow-up of SEMIS I was requested by the Greater Mekong Subregion (GMS) countries at the Fourth Meeting of the GMS Working Group on Environment held in Hanoi in March 1998. GMS countries are Cambodia, Yunan province of the People's Republic of China, Lao People's Democratic Republic, Myanmar, Thailand, and Vietnam. The GMS Ministerial Meeting held in Manila in September 1998 endorsed the request. The project was approved by ADB on 29 December 1999.

The objective of the Technical Assistance is to build upon the achievements of SEMIS I which include: (1) a defined core dataset, (2) a conceptual spatial database design, and (3) technical capacity for the exchange of data. The overall goal of SEMIS II is to help GMS governments to make informed decisions regarding sustainable development through integrated economic and environmental planning.

In January 2001, the United States Geological Survey (USGS) was asked to provide expertise for the SEMIS II project in terms of geographic information systems and spatial databases. Mr. Michael Starbuck, USGS, spent 8 weeks at the Center from January 29 to March 23, 2001. Terms of reference included: (1) Review and analyze available information at the Center to determine the usefulness and relevance for operational level development planning for the GMS Hotspot areas, (2) Review the identified gaps and suggest additional data/information needs that may be required for the operational level planning purposes, (3) Suggest the appropriate scale and format for GMS spatial data used for operational planning, and (4) Review the data collection and management guidelines prepared by the SEMIS I team and suggest revision and refinements. This report summarizes items 1 through 3, with a separate report, entitled "Draft Data Collection and Management Guidelines", covering item 4.

Due to time constraints and scheduling difficulties, the following items from the terms of reference were not completed: (1) Assist in analyzing temporal land use/land cover changes in the GMS countries using available software packages, (2) Assist in organizing national training seminars to be conducted in Kumming-Yunnan, PRC, Vientiane, Lao, PDR, Phnom Penh, Cambodia, Hanoi, Vietnam, and Bangkok, Thailand. At these seminars, make presentations on collection and processing of data/information, and (3) Assist in evaluating the status of data processing in the GMS countries and suggest personnel and material (hardware/software) needs to establish or strengthen to data centers, as appropriate. The SEMIS II project is not advanced sufficiently to yet schedule national training seminars. A trip was made to Phnom Penh, 1- 3 March 2001, to attend a Strategic Environmental Framework project meeting, and to visit the Mekong River Commission office.

2.0 Background and objectives

The overall goal of the SEMIS II project is to help the GMS governments to make informed decisions regarding sustainable development through integrated economic and environmental planning. A key component of informed decision making is having access to reliable information, including spatial data. The GMS countries, because of their geographic locations, share common environmental problems and therefore have a common need to share environmental information in a timely manner. SEMIS I, the Technical Assistance project that was the precursor to SEMIS II, established the groundwork for the sharing of information on environmental and natural resources issues.

A related on-going Technical Assistance project is the Strategic Environment Framework (SEF) for the GMS (TA No. 5783). This project has similar goals of integrating economic and environmental planning and is also collecting geospatial data for selected areas in the GMS.

SEMIS II aims to build upon the achievements of SEMIS I by undertaking pilot demonstration projects, further developing subregional/national databases, reviewing the current mechanism for collection of identified core data, and defining the best approach to performing data collection, storage, manipulation, and transmission.

The key objectives of SEMIS II include:

1) Assess the availability of useful and relevant data for planning purposes;

2) Increase and strengthen the capacity of national governments to collect and process the information/data;

3) Increase the capacity of national governments to make informed decisions regarding development investments relating to sustainable use of natural resources;

4) Enhance the ability of GMS national governments to conduct integrated economic and environmental planning with relevant data; and

5) Conduct, store, manipulate and share actual integrated planning information using the data collected in pilot projects for some "Hotspot" areas, such as those identified in TA 5783-REG: Strategic Environmental Framework for the GMS.

The following are the specific outputs planned for the SEMIS II project (see the SEMIS II Implementation Plan for detailed information):

1) Available data and data gaps,
2) Integrated economic and environmental planning procedures and background papers,
3) Capacity building plan on hardware/software support and training needed,
4) Guidelines for data collection and data management,
5) Hardware and software support,

6) Internship-cum-training for six national coordinators,
7) Enhancement of a sub-regional network,
8) Study tour of national coordinators/finance personnel,
9) "Hot spot" database (1:50K),
10) Baseline data of GMS (1:250-500K),
11) Case studies,
12) Sub-regional and national training/seminars/workshops and study tours,
13) Project reporting and management.

2.1 Objective of this activity

This report addresses the activities supporting the SEMIS II Objective 1 (Assess the availability of useful and relevant data for planning purposes), and specifically Output 1 – Available Data and Data Gaps. The product is this report on the availability and accessibility of data and data gaps.

2.2 Appro ach

The following steps were used in determining the availability and accessibility of data and data gaps:

1) Review outputs of earlier projects (SEMIS I and SEF);
2) Identification of core datasets required for analysis;
3) Inventory of available datasets using the Data Catalogue; and
4) Identification of data gaps.

3.0 Review of earlier projects

This section summarizes the main points of the earlier projects and presents recommendations on ways to improve upon the results.

3.1 Subregional Environmental Monitoring and Information System (SEMIS I)

SEMIS I is the Technical Assistance project financed by the Asian Development Bank (T.A. No. 5622-REG), which was the precursor to the current SEMIS II project.
SEMIS I was approved in January 1995 and was completed in November 1999. There were 17 different ouputs from the SEMIS I project, with responsibility for individual outputs varying between the Center, the ADB project team headed by Roche International of Quebec, and the Mekong River Commission. This review will focus primarily on the following outputs: (1) core dataset definition, (2) spatial database design, (3) data standards, (4) metadata standards, and (5) catalogue of data holdings.

3.1.1 Output 1 – Core Dataset Definition

A core dataset is defined as,

"Identification and description of the 'core' or 'minimum' set of spatial information needed to support national and subregional environmental assessment, decision making, and environmental reporting."

Another definition used is,

"The basic, frequently required data necessary for the range of environmental decisions which will arise in subsequent years."

The core dataset definition resulting from the SEMIS I activities was designed to support a wide range of national and subregional environmental decision making and analysis. By definition, the core data sets would be collected and integrated across the GMS countries according to established guidelines and standards, thus making them useful to a variety of decision makers in a timely fashion.

The SEMIS I activities which lead to the core dataset definition were:

- review of previous studies of relevance in the region
- preparation of a draft set of core datasets
- consultation with the six countries, and
- consolidation of the findings into the final core dataset definition.

The following are the 13 core datasets as defined by the SEMIS I team:

1. Infrastructure
2. Soil Class
3. Vegetation Cover
4. Air Quality Measurements
5. Demography
6. Climate Zonation
7. Administrative Boundaries
8. Topography (and Hydrography)
9. Land Use
10. Geology
11. Major Harvesting Activities
12. Water Quality Measurements
13. Soil Analysis Samples

Topography and hydrography

While definitely related, topography and hydrography, in my opinion should be treated as separate data layers. These data layers can be derived independently using modern technology (i.e. LIDAR and interferometry) and many users may wish to use one or the other independently.

Data collection issues

Infrastructure, administrative boundaries, topography, and to some extent, vegetative cover and land use, are all data layers which historically can be found on traditional topographic maps. Initial data collection efforts can concentrate on the digitization of existing map features from paper maps. A shortfall to this approach, however, is that you are constrained by the interpretations performed by the original mapmaker. Questions that must be considered include: what features did they actually collect, and what classification scheme did they use and does it match your own? The best approach is to clearly document the map features collected and the classification scheme used, preferably in a standard metadata format, so that subsequent users know exactly what kind of information is present in the data sets.

The standards for the core datasets as defined by the SEMIS I project have established guidelines for feature collection (at least in the case of infrastructure), which have not considered the original data source. For example, the attributes available to the map feature "Main Road", include the following possibilities: international, national, and secondary. This is a very generic, simple classification scheme. However, the most likely data source for these map features will be the national 1:250,000 scale topographic maps. Roads on these maps are classified according to surface type and number of lanes. There is no clear and reliable translation from that scheme to the other. Determination of whether a road is national or international may be very subjective. Besides taking advantage of the existing map symbolization (and classification scheme), a scheme based on observable characteristics is preferable to one based on interpretation of a features' use.

Framework perspective

Another perspective to consider is that from the Federal Geographic Data Commission (FGDC). The FGDC has identified a set of data layers it has named the "Framework". These data layers are the common themes needed by most data users and include transportation, hydrography, geodetic control, digital imagery, boundaries, elevation, and cadastral layers. To make the use and distribution of framework data easier, the FGDC is proposing certain technical, operational, and institutional contexts, such as a feature-based data model, permanent feature identification codes, references to datums, and metadata.

While the core data list proposed by SEMIS I is more inclusive, having been designed with a more specific user base in mind, many of the concepts behind the framework design would apply and it may be worth examining in more detail. See the FGDC Framework website at http://www.fgdc.gov/framework/framework.html.

Additional core data layers

Another data layer usually listed as a core dataset is imagery. Imagery can be in many forms, including digitized air photos, digital orthoimagery (airphotos processed to remove distortions due to terrain relief), satellite imagery, and scanned paper maps. Landsat 7 ETM+ scenes are relatively low cost and have no re-distribution restrictions (the data is not copyrighted). Landsat scenes are an excellent data source for a variety of map features, from basic infrastructure to land cover.

Scanned paper maps are an inexpensive alternative to performing vector digitizing of topographic maps. The original paper maps can be scanned wherever a service provider has large format scanning capability and then the raster image can be georeferenced using standard geoprocessing software. The result is a georeferenced image of the topographic map that can be used for a number of applications, ranging from performing heads-up digitizing of map features to serving as an inexpensive base map for other project data.

Slope and aspect are two data layers often critical to certain kinds of analysis. While these layers are derivable from the elevation dataset, some consideration might be given to having these layers pre-existing, for those users who may not have the capacity to create them themselves.

While the data layer, Water Quality Measurements, is one of the core datasets, environmental impact analysis could benefit from more detailed surface and subsurface hydrologic data. Currently only water sample sites are specified, listing various characteristics of the sample (pH, conductivity, chemical analysis, etc.). Additional measurements like depth to groundwater, surface water hydrographs, basin delineation, and subsurface flows would be valuable to performing analysis of potential impacts of development projects.

Responsibility for data layers

Creating lists of core, required data sets is a useful exercise to perform. It gets the stakeholders thinking about what kinds of datasets they may regularly require in their decision-making activities. However, for these data layers to become more than "wish lists" on paper, responsibility must be assumed for their creation, maintenance, and distribution. One of the weaknesses of the SEMIS I output concerning core datasets is the lack of a strong assignment of responsibility for the individual datasets. One approach to this is to get the appropriate agency or group most aligned with a particular data layer to sign on as the data supplier for that data set. For example, the government agency responsible for forestry may be the best equipped to provide the vegetation cover data layer. The transportation department would be most likely to provide at least some of the infrastructure layer. Of course, this will not work in many cases, and in the GMS countries, only a handful of agencies may be capable of creating and distributing geospatial data. This brings us back to the basic underlying premise for the SEMIS projects – establish among the GMS countries a mechanism for the efficient exchange of geospatial data for timely economic and environmental planning.

3.1.2 Output 2 – Spatial database design

From the SEMIS I implementation document – A conceptual level spatial database design for a hierarchical subregion wide GIS to support national and sub-regional environmental

assessment, decision making and environmental reporting. The design should consider a distributed system of inter-linked spatial databases at the national level (target scale 1:50K) which can be integrated into a sub-regional GIS (target scale 1:250K). The primary function of the database is to manage and analyse the Core Dataset defined in Output 1.

The SEMIS I team used a list of criteria to help govern the development of the conceptual database design. The database should be:

- Decentralized,
- Hierarchical,
- Spatially based,
- Expandable and flexible,
- Easy to use and maintain,
- Built using appropriate technology,
- Compatible with UNEP State of the Environment Database, and
- Compatible with Existing Subregional Databases.

The conceptual database design is reasonable and logical in its structure. A national hub in each country acts to obtain and exchange data with other countries using the standard exchange formats. The subregional hub, probably an international agency such as UNEP or MRC, will link to the national hubs and other international agencies to exchange data.

A key component of the design is the concept of decentralization and that there should be no duplication of data. At the national level, a number of different agencies will hold various components of the core datasets. Data exchange would occur through the hub agencies.

The SEMIS functional design discusses the kinds of functions that will be required of a SEMIS system to support environmental monitoring and reporting:

- Manage Core Datasets
- Manage Non-core Datasets
- Produce SEMIS Outputs
- Manage Auxillary Datasets
- Manage Dataset Availability
- Manage Dataset Exchange
- Convert Existing GIS Datasets

The subsystem “Produce SEMIS Outputs” contains major subcomponents:

- Spatial Data Management
- Spatial Data Analysis
- Spatial Output and Display
- Non-spatial Analysis

To provide these kinds of functions, a full function GIS is required. The SEMIS report indicates not every participating agency will need to have this functionality and that it may be best initially to concentrate GIS capability in one or a few centers in each country. For the

majority of users, desktop computers running the latest version of Arcview will provide this functionality.

The FGDC is developing a metadata distribution mechanism that may have implications for data distribution in the GMS. It is called a Data Clearinghouse and is a decentralized system of servers located on the Internet that contains descriptions of available digital spatial data (metadata). The fundamental goal of the Clearinghouse is to provide access to digital spatial data through metadata. For more information, see the website at http://www.fgdc.gov/clearinghouse/clearinghouse.html

3.1.3 Output 3 – Data standard for information to be held in national and sub-regional databases

From the SEMIS I implementation document: Following from the Core Dataset Definition (Output 1) these standards and guidelines are to ensure the feasibility of data exchange and integration. They will include preferred classification systems for core data items, spatial data frameworks and recording standards, and standard data interchange formats. They are not meant to extend to standards for data collection, measurement methods, data coding, or quality control procedures.

A great deal of work has gone into developing a detailed data standard for the core datasets. The general approach taken was to rely on the organization of expertise for a given data layer. For example, the standards adopted for the soil class data set are those defined by FAO in the Global and National Soils and Terrain Digital Databases (SOTER) Procedures Manual (FAO, 1993).

The attribute scheme for the infrastructure datasets is a bit rudimentary, and as discussed earlier, the relationship between the desired classification scheme and those already in use may present a problem.

The main point for discussion here is whether the standards are, or will be, used. There is little evidence that the data standards are currently being used, even by the international agencies in the region. Perhaps as the national agencies begin their data collection activities in earnest, they will follow the proposed standards. More likely, however, is that as the individual agencies create their own data sets, they will devise a scheme of their own design, emphasizing the features and attributes that are important to them. Adhering to standards is a difficult process, especially amongst agencies and countries that may not see doing so in their best interest.

Spatial data exchange format standard

Internationally, a great deal of work is going into devising data exchange formats. Of particular note are the USGS Spatial Data Transfer Standard (SDTS), and the efforts of ISO Technical Committee 211. It was the recommendation of the SEMIS I group to use the Arc/INFO Export Format for vector GIS data as an interim standard for data exchange. This was to be in effect until the ISO 211 standard was complete. As GMS countries will be using commercial GIS software packages, it makes sense to use an existing exchange format supported by the software. Arc/INFO Export Format will be suitable in the vast majority of cases.

The ASCII coded headerless files have been selected as the interchange format for raster files. This is suitable as a basic common format, but in practice, other widely used raster formats will probably be used more often, such as TIFF and GeoTIFF.

3.1.4 Output 4 – Metadata standards for information to be held in national and sub-regional databases

The agreed interim standard format for metadata is that developed by UNEP-GRID. This is already used in the sub-region and is compatible with the major metadatabases of NASA, CIESIN, WCMC and others. An existing metadata entry tool, a Microsoft Access-based software package developed by UNEP, is available for creation of new metadata files.

This standard is a good choice for metadata, as it is already accepted and based on international standards. Metadata creation is a difficult and time-consuming task. If more complicated metadata standards were required, it is possible that little metadata would actually be created. The practice of creating metadata should be reinforced at every opportunity. When the training for the national coordinators is conducted, a standard metadata file should be created for each spatial data file created by the trainees.

3.1.5 Output 5 – Catalogue of existing environmental and natural resources data holdings amongst countries of the GMS

The Center maintains a catalogue of datasets held by agencies in the Asian and Pacific regions. The catalogue is updated every 6 months. The issue reviewed was dated October, 2000. Eleven institutions contribute information concerning data holdings. They are: International Center for Integrated Mountain Development (ICIMOD) in Nepal; Mekong River Commission (MRC) in Cambodia; South Pacific Regional Environment Programme (SPREP) in Western Samoa; UNEP-GRID in Bangkok; Landcare Research in New Zealand; Ministry of Environment in Cambodia; International Irrigation Management Institute (IIMI) in Pakistan; National Mapping and Research Information Authority (NAMRIA) in Philippines; Department of Environmental Quality Promotion (DEQP) in Thailand; National Environment Agency (NEA) in Vietnam; and SENRIC Project (SACEP) in Sri Lanka.

While the Center maintains the list, only those datasets listed under the UNEP-GRID data holdings may be obtained from that office. For datasets of other agencies, one must contact that agency directly using the supplied contact information. Datasets can be distributed unconditionally, distributed with source approval, or made available for in-house use only, depending on the particular dataset. This service was originally provided free of charge, but a $50 per data request fee is now applied. The data listing is also available from the UNEP website at http://www.eapap.unep.org/fs-datacat.html.

The data listing gives the following information for each dataset: code, title, type(vector or raster), general location, date, scale, and size of digital file. This listing is a wonderful resource for those looking for spatial data. From the website version, users can access metadata listings for a selected subset of the datasets listed. Metadata files need to be created for all the datasets listed in the catalogue. Actual data requests seem to be very infrequent, with approximately 2 requests coming in to the Center in the past year. When a data request is received, it must be made clear to the user that a fee is required. Then the appropriate files are found on the stored CD-ROMs and copied to the medium of choice, either 3.5 inch

diskette, or CD-ROM. Occasionally the user will request some additional processing, such as data subsetting, or conversion to other coordinate systems.

Hopefully, in the future, additional agencies will participate in this data listing. Currently, of the GMS countries, Lao, Myanmar, and China are not represented in the Data Catalogue. There are also undoubtedly many more agencies that have geospatial datasets that are not currently participating in the catalogue. Also, the majority of the datasets listed in the catalogue are small scale, at 1:250,000 or smaller. Most agencies are reluctant to make larger scale datasets available outside their department, let alone to the general public or other countries. Perhaps some kind of incentive can be devised to encourage more participation and the release of larger scale datasets.

3.2 Strategic Environment Framework for the GMS (SEF project) TA No. 5783

The overall objective of the SEF Project is to promote a better understanding of environmental and social impacts of planned development in the GMS. In particular, emphasis is given to the energy/water resource and transportation sectors of the ADB's GMS Programme. A key component is to help ensure that environmental and social aspects are considered at an earlier stage in the planning process than currently takes place.

SEF outputs include:

- A report that will provide a framework of operational, policy, and institutional recommendations designed to better ensure the environmental and social sustainability of economic development;
- A list of recommended GMS Technical Assistance (TA) projects and environmental investments;
- A set of maps and GIS databases on baseline conditions in the region;
- A set of maps and GIS databases on GMS environment-development "Hotspots" and Highly Valued Areas;
- GMS development scenarios;
- A GIS-based GMS *Early Warning and Information System (EWIS)*.

3.2.1 Early Warning and Information System (EWIS)

The EWIS is an interactive software tool, built upon the Arcview software package from Environmental Systems Research Institute (ESRI). The tool requires Arcview to already be installed on the system to work. Through the use of modified menus and tools in Arcview's graphical user interface, the EWIS presents the user with an easy way of viewing and performing simple queries on GMS datasets. The system provides information to the user on Highly Valued Areas, the environmental and social status of the GMS, development plans in the GMS, and priority Hotspots. It also allows the user to see, in a general way, possible impacts to proposed development projects. The user can draw a proposed new highway, or the location of a proposed dam, and the system will automatically calculate what other data layers are affected.

Most of the geospatial data associated with the EWIS is at the 1:1 million scale. Table 1 lists most of the datasets accessed by the EWIS. This level of detail is appropriate for viewing features and relationships at the regional level, but the user must keep in mind the constraints imposed by such a relatively small-scale database. Arcview will allow the user to zoom in

almost indefinitely, but the appropriate display scale of the data can be quickly surpassed, perhaps leading to incorrect conclusions.

In addition to the regional datasets, the SEF project is collecting higher resolution data for specific areas termed "hotspots". These are areas determined to represent the conflict between economic development and environmental and social goals in the GMS. Five hotspot areas have been delineated in the GMS region. The detailed hotspot analysis will be undertaken at the 1:250,000 scale using data provided by the MRC, as well as information collected by the SEF team and by UNEP.

File Name	Data Description	Scale	Source
GMS_bnd.shp	Country Boundaries for the Greater Mekong Subregion	1:1M	UNEP
Asia_bnd.shp	Regional Boundaries for other countries in Asia	1:5M	ESRI
Gms-riv.shp	Greater Mekong River System	1:1M	UNEP
Gms-wet.shp	Waterbodies in the GMS	1:1M	MRC, UNEP,China
Hotspot.shp	SEF Priority Hotspots	1:1M	SEF
Pov_cam.shp	Poverty data for Cambodia	-	MoE, Cambodia
Pov_lao.shp	Poverty data for Lao PDR	-	STEA, Nat. Stat.
Pov_mya.shp	Poverty data for Myanmar	-	For. Dept. Myanmar
Pov_thai.shp	Poverty data for Thailand	-	UNEP
Pov_viet.shp	Poverty data for Vietnam	-	-
Pov_yun.shp	Poverty data for Yunnan	1:1M	YEPB
Gms_hva.shp	Highly Valued Areas in the GMS	1:1M	SEF
Mrc_forest.shp	MRC forest cover for the lower Mekong Basin	1:250K	MRC
Wwf_bio.shp	WWF for Nature Biodiversity Priorities for the forests	1:1M	WWF
Mrc_wet.shp	MRC data for wetlands in the lower Mekong Basin	1:250K	MRC
Unep_land.shp	UNEP Landcover data for SE Asia	1:1M	UNEP
Gms_pa.shp	Protected areas in the GMS	1:1M	UNEP
Gms_road.shp	Roads in the GMS	1:1M	UNEP
Gms_road_prj.shp	Proposed ADB road projects	1:1M	ADB
Gms_dam.shp	Current Dam locations	1:1M	UNEP
Gms_city.shp	Locations of cities	1:1M	UNEP
Gms_prov.shp	Provincial boundaries in the GMS	1:1M	UNEP
Gms_dist.shp	District and county boundaries in the GMS	1:250K	MRC

Table 1 – Datasets used by the EWIS

4.0 Review of Available Digital Geospatial Datasets

The primary document used for this review was the Data Catalogue, published by Center, October 2000. The procedure used for review was to examine the list of Core Datasets and then populate a table, indicating a dataset file name where it was determined that it satisfied the Core Dataset definition. In this way, the tables in appendix 1 were constructed, showing where datasets were available, and where they were not. A limitation to this method was that the actual datasets were not examined, only the short description in the Data Catalogue, or in some cases, an actual metadata listing, was all that was available for determining suitable matches to the Core Data list.

A series of asterisks (*) in the file code field indicates no match was found in the Data Catalogue for this entity. No tables were generated for China or Myanmar, as very little data at the larger scales is listed in the Catalogue for these countries.

More datasets exist than what is listed in the Data Catalogue, especially for Thailand. These additional datasets need to be documented as to their existence, either in the Data Catalogue, or in a separate listing.

4.1 Map accuracy issues

Throughout the SEMIS I reports and other related documents, there has been little mention of map and data accuracies. It is usually desirable to determine the level of accuracy of a given map or series of maps. There are many kinds of accuracy assessment, including how complete a map is (i.e. for a road map, are all the existing roads depicted on the map?), the attribution (does that road have the correct identifier?), to positional accuracy (is the road shown in the right place). There is also absolute and relative positional accuracy (not only is the road in the right place, but is that house on the correct side of the road?). Accuracy assessment can vary from a rigorous examination of each dataset, to a statistical sampling of selected datasets. Generally you can rely on the process (techniques and methods used to create the map) to retain the desired quality and accuracy in the product (final dataset), but occasionally verification is desired to be certain the required accuracy is indeed being achieved. I recommend creating some data validation plan that includes assessment of map accuracy and applying it to datasets generated by the project. Appendix 2 discusses how U.S. National Map Accuracy is defined and measured.

4.2 Map projections and coordinate systems

Some mention needs to be made of which map projections, datums, and coordinate systems are to be used for the SEMIS II project. As most of the map datasets are being digitized from existing map sources, they will already have some projection and be cast on a particular datum. Will all datasets be transformed into a common set of datum, projection and coordinate system? Geographic coordinates are a good general purpose coordinate system, but the user often must transform the dataset into a projected map before they can easily make measurements. Arcview will allow the dataset to remain in geographic coordinates and automatically perform a transformation to a specified projection so that distance and area measurements can be taken. In Arcview, if some of your data is in geographic, all other data used in that view must also be in geographic for proper registration to occur.

4.3 Recommended data scales/resolutions

The key to establishing recommended data scales is to examine the kinds of measurements or analysis one hopes to perform with the data. To use the EWIS from the SEF project as an example, its primary objective is to provide users with an overview of various developmental and environmental aspects of the GMS region and allow for identification of potential impacts of proposed developments. Most of the data layers being used by the EWIS are at the 1:1 million scale. If we allow that the datasets meet a level of horizontal accuracy commensurate with the U.S. National Map Accuracy Standard (which may not be the case), a well-defined point in the dataset can only be expected to be within approximately 500 meters of its true horizontal position. This is only a rough estimate and will vary according to a number of factors. For the EWIS application, this is an acceptable value, as users will not (hopefully) be trying to make measurements at a resolution finer than this. The horizontal accuracy of curvilinear features will probably not be as reliable as well-defined points, and when you overlay multiple layers, you must account for the possible accumulation of errors. Scale also implies a certain level of content detail. A road layer created at the 1:50,000 scale

will have more detailed roads (both in showing smaller roads and more detail of the linework) than a road map at a scale of 1:250,000.

For analysis and planning purposes at the regional scale (the extent being across several GMS countries), a map scale from 1:500,000 to 1:1 million is reasonable. For more detailed analysis at the level of area hotspots (extents of about 300 kilometers), a map scale of 1:250,000 or larger is desirable. This scale implies a maximum positional error of well-defined points of about 130 meters. Again, these positional error calculations are rough estimates, and there are many variables that should be considered.

4.4 Data accessibility

For the datasets listed in the Data Catalogue, access can be considered fair to good. According to the documentation associated with the Catalogue, the user simply mails a data request form to the appropriate agency indicating the datasets desired, distribution media, and any special requirements. Depending on the data requested (and possibly on the agency being dealt with), the process of requesting and then receiving a particular dataset may take anywhere from a few weeks to several months. An added complication is the fact that some agencies are charging distribution fees and are not advertising this fact. Also, the dataset description should be more detailed, so that the users can be sure of getting the datasets they require.

While this dataset distribution scheme is not as convenient as online Internet access, it is better than nothing and has the added convenience of bringing together datasets from a variety of agencies in the region. Possible weak links include the need for continuity of communication between a coordinator at the UNEP office and contacts at the respective agencies and the establishment of efficient local data archives at each agency. As the technological infrastructure improves in the region, migration to some Internet-based data distribution will be easier because of the current scheme.

5.0 Data gaps

In reviewing the available data compared to the desired core datasets, several data gaps are noticed. Consistent among the various GMS countries is the apparent lack of certain infrastructure layers, such as electric transmission lines, pipelines, dams, ports, and airports. The SEF project does use a layer in its EWIS software for dams in the GMS region, but the source scale is 1:1 million. Given adequate topographic maps, these data layers could easily be collected in a minimal amount of time.

Another layer typically not available is the air quality measurements. These would most likely be large amounts of tabular data linked to a point location. If the source for the tabular data can be obtained, creation of spatial data layers would be fairly easy. The same is true for the demography and water quality data layers.

The last group of data layers consistently not available are the Major Harvesting Activities datasets, including agriculture, forestry, mining, and fisheries. These data layers are defined primarily as attributed polygons at the province or district level. If the tabular data can be obtained, linking it to existing provincial or district spatial data would be fairly straightforward.

6.0 Conclusion / Recommendations

This report addresses Output 1 of the SEMIS II project – Available data and data gaps. The approach used was to review the outputs of the SEMIS I and SEF projects, identify those core datasets required for analysis, inventory available datasets using the Data Catalogue, and identify the data gaps. Areas of apparent data gaps include:

- Infrastructure (electric transmission lines, pipelines, dams, ports, and airports),
- Air quality measurements,
- Demography,
- Water quality measurements, and
- Major harvesting activities.

The core dataset list, data standards, and metadata standards as defined by the SEMIS I team were examined and the following are suggested improvements:

- Separate topography and hydrography into separate layers or themes;
- Consider feature classification schemes and tailor to meet requirements and existing data sources;
- Consider applying a framework approach to the SEMIS II database activities;
- Possible additional core data layers include:
    - Remotely sensed imagery,
    - Raster topographic maps,
    - Slope and aspect, and
    - Detailed hydrologic information;
- Assign responsibility for data layers;
- Use an easily populated metadata standard;
- Arc/Info export files are a suitable interchange format;
- Devise and use a data validation and accuracy assessment plan;
- Address the issues of multiple datums, projections, and coordinate systems;
- Clearly advertise data distribution fees;
- Strengthen procedures for maintenance of the database and the data catalogue; and
- Generate metadata files for all data catalogue listings.

Entity	File Code	Scale	Date	Comments
Infrastructure				
Main Road	ALL0008	250,000		Provincial datasets from OEPP
Railway	**********			(ALL0008 refers to CD-ROM)
ElectricTransmission Line	**********			
Pipeline	**********			
Dam	**********			
Port	**********			
Airport	**********			
Soil Class				
Soil Unit	ALL0008	250,000		Provincial datasets from OEPP
Vegetation Cover				
Land Cover Unit	THA 0028	1,500,000	1991	Forest-Nonforest of Thailand
	THA 0033	1,000,000	85-86	Land Cover map of Thailand (85-86)
	THA 0034	1,000,000	92-93	Land Cover map of Thailand (92-93)
	TH-WL 100	250,000	1997	Wetlands in the Mekong Corridor, Thailand
	ALL0008	250,000		Provincial datasets from OEPP
Air Quality Measurements				
Air Quality Observation	**********			
Demography				
Demographic Unit	**********			
Climate Zonation				
Agro-climatic Zone	**********			
Administrative Bdys				
Administrative Unit	ALL0008	250,000		Provincial datasets from OEPP
	TH-AM001-017	50,000	1969	Administrative maps of NE Thailand from MRC
Topography				
Elevation				MRC DTM's ?
	ALL0008	250,000		Provincial datasets from OEPP
Water Boundary / Body	ALL0008	250,000		Provincial datasets from OEPP
Land Use				
Land Use Unit	ALL0008	250,000		Provincial datasets from OEPP
Geology				
Geological Unit	TH-GL008-100	250,000		Geological maps of NE Thailand from MRC
Major Harvesting Activities				
Agriculture	**********			
Forestry	**********			
Mining Location	**********			
Fisheries	**********			
Water Quality Meas.	**********			
Soil Analysis Samples	**********			

Appendix – 1: Thailand Available Datasets

Entity	File Code	Scale	Date	Comments
Infrastructure				
Main Road	NEA 3	250,000	1997	Topography map of vietnam
Railway	NEA 3	250,000	1997	Topography map of vietnam
ElectricTransmission Line	**********			
Pipeline	**********			
Dam	**********			
Port	**********			
Airport	**********			
Soil Class				
Soil Unit	VN-SL 001	250,000	1989	Soil Map of Mekong delta
	VN-SL 002	250,000	1989	Soil Map of Mekong delta (raster)
	NEA 30	250,000	1997	Soil Map of north -west area of Vietnam
	NEA 31	250,000	1994	Soil Map of Mekong river delta
	NEA 32	250,000	1990	Soil Map of Centro-highland Tay nguyen
	NEA 33	250,000	1994	Soil Map of Centro-coastal area of Vietnam
	NEA 34	250,000	1994	Soil Map of South-East area of Vietnam
	NEA 25	1,000,000	1996	Soil Map of Vietnam - FAO/UNESCO classifi.
	NEA 26	250,000	1994	Soil Map of Vietnam - former Soviet Union class.
	NEA 27	250,000	1997	Soil Map,NW area of Vietnam - FAO/UNESCO
	NEA 29	250,000	1994	Soil Map of Red river delta
Vegetative Cover				
Land Cover Unit	VIE 0005	1,000,000	85-86	Land Cover map of Vietnam 85-86
	VIE 0006	1,000,000	92-93	Land Cover map of Vietnam 92-93
	NEA 3	250,000	1997	Topography map of Vietnam
	NEA 5	500,000	1997	Database of forestry cover of Vietnam (1943)
	NEA6	500,000	1997	Database of forestry cover of Vietnam (1983)
	NEA 7	500,000	1997	Database of forestry cover of Vietnam (1995)
	NEA 38	250,000	1995	Ecological Map of Red river delta
	NEA 40	250,000	1987	Ecological Map of Mekong river delta of Vietnam
Air Quality Meas.				
Air Quality Observation	VN-MT 001	500,000	1993	Meteo-monitoring Stations ?
Demography				
Demographic Unit	NEA 4	100,000	1997	Demography map of Vietnam
Climate Zonation				
Agro-climatic Zone	NEA 50	1,000,000	1997	Climate map of Vietnam (precip and temp)

Appendix – 2: Vietnam Available Datasets

Administrative Bdys				
Administrative Unit	VN-AM 001	250,000	1993	Administrative Boundary in Mekong delta (raster)
	NEA 1	250,000	1997	Administrative map of Vietnam
	NEA 2	1,000,000	1997	Administrative map of Vietnam
	NEA 3	250,000	1997	Topography map of Vietnam
Management Areas	NEA 8	100,000	1997	Database of Protected Areas, NP of Vietnam
Topography				
Elevation				MRC DTM's ?
	VIE 0004	5 minute		Derived from ETOPO5
Water Boundary / Body	VN-IN 001	250,000	1986	Inundation Map of Mekong delta (raster)
	LM DN 005	250,000	1956	Drainage map - Ca Mau
	LM-DN 006	250,000	1956	Drainage Map - Vinh Loi
	LM-DN 007	250,000	1967	Drainage Map - Cai Nuoc
	LM-DN 026	250,000	1954	Drainage Map - Hoai Nhon
	LM DN 004	250,000	1955	Drainage Map - Saigon
Land Use				
Land Use Unit	VN-LU 001	250,000	1993	Landuse map of Mekong delta (raster)
	NEA 20	1,000,000	1992	Landuse Map of Vietnam
	NEA 21	250,000	1992	Landuse Map of Vietnam
	NEA 22	750,000	1992	Landuse Map of Vietnam
	NEA 23	1,000,000	1995	Land Map Unit (Land evaluation map) of Vietnam
	NEA 28	250,000	1994	Land Map Unit (Land evaluation map) of Vietnam
Geology				
Geological Unit	NEA 9	500,000	1980	Geology map of Vietnam
	NEA 10	200,000	1995	Geology map of Vietnam
	NEA 11	200,000	1995	Geology map of Northern part of Vietnam
	NEA 12	500,000	1980	Hydrogeology map of Vietnam
	NEA 24	1,000,000		Geomorphology map of Vietnam
Major Harvesting Act.				
Agriculture	**********			
Forestry	**********			
Mining Location	**********			
Fisheries	**********			
Water Quality Meas.	**********			
Soil Analysis	**********			

Appendix – 2: Vietnam Available Datasets (cont.)

Entity	File Code	Scale	Date	Comments
Infrastructure				
Main Road	CAM0018	1,000,000	1988	Roads, Railroads map of Cambodia
Railway	CAM0018	1,000,000	1988	Roads, Railroads map of Cambodia
ElectricTransmission Line	**********			
Pipeline	**********			
Dam	**********			
Port	**********			
Airport	**********			
Soil Class				
Soil Unit	CA-SO 001	500,000	1986	Soil Map of Cambodia
Vegetation Cover				
Land Cover Unit	CAM 0041	1,000,000	1971	Vegetation Map of Cambodia
	CA-VE 001	1,000,000	1971	Vegetation Map of Cambodia (same as above?)
	CA-LU 1010	250,000	1993	Landuse land cover map of Cambodia
	CAM 0044	1,000,000	85-86	Land Cover m ap of Cambodia (85-86)
	CAM 0045	1,000,000	90-91	Land Cover map of Cambodia (90-91)
Air Quality Meas.				
Air Quality Observation	**********			
Demography				
Demographic Unit	**********			
Climate Zonation				
Agro-climatic Zone	CAM 0017	2,000,000	1968	Climatic Zones map of Cambodia
Administrative Bdys.				
Administrative Unit	CA-AM 100	500,000		Cambodia Provincial Boundaries
	CA-AM 101	500,000		Districts Boundaries
Management Areas	**********			
Topography				
Elevation				MRC DTM's ?
	CAM 0007	5 minute		Derived from ETOPO5
Water Boundary / Body	CAM0019	1,000,000	1988	Rivers, Lakes, Islands map of Cambodia
	CAM 0021	2,000,000	1968	Drainage (Flooded area) map of Cambodia
	LM-DN 001	250,000	1954	Drainage Map - Phnom Penh
	LM-DN 002	250,000	1955	Drainage Map - Prey Veng
	LM-DN 003	250,000	1954	Drainage Map - Long Xuyen
	LM DN 004	250,000	1955	Drainage Map - Saigon
	LM DN 018	250,000	1954	Drainage Map - Battambang
	LM DN 019	250,000	1954	Drainage Map - Siemreap
	LM DN 020	250,000	1954	Drainage Map - Stung Treng
	LM DN 021	250,000	1954	Drainage Map - Veune Sai
	LM DN 022	250,000	1954	Drainage Map - Chanthaburi
	LM DN 023	250,000	1954	Drainage Map - Pursat
	LM DN 024	250,000	1954	Drainage Map - Kratie
	LM DN 025	250,000	1954	Drainage Map - Sre Khtum
	CA-IN 001	500,000	1982	Inundation Map of Cambodia
	CA-WB 100	250,000	1993	Drainage and open water bodies of Cambodia

Appendix – 3: Cambodia Available Datasets

nd Use				
nd Use Unit	CAM 0038	2,000,000	88-89	Reconnaissance Landuse Map of Cambodia
	CA-LU 100	250,000	1991	Landuse Map of Cambodia
	CA-LU 001	250,000		Landuse map of Tonle Sap area
	CA-LU 002	250,000		Landuse map of Sambor area
	CA-LU 003	250,000		Landuse map of Stung Treng area
	CA-LU 101	250,000	1993	Landuse land cover map of Cambodia
ology				
eological Unit	LM-GL 100	1,000,000	1988	Geology of Cambodia, Laos, Vietnam
ajor Harvesting Act.				
riculture	**********			
restry	**********			
ning Location	**********			
sheries	**********			
ater Quality Meas.	**********			
il Analysis Samples	**********			

Appendix – 3: Cambodia Available Datasets (cont.)

Entity	File Code	Scale	Date	Comments
Infrastructure				
Main Road	LAO 0015	1,000,000	1988	Roads, Railroads map of Laos
Railway	LAO 0015	1,000,000	1988	Roads, Railroads map of Laos
ElectricTransmission Line	**********			
Pipeline	**********			
Dam	**********			
Port	**********			
Airport	**********			
Soil Class				
Soil Unit	**********			
Vegetation Cover				
Land Cover Unit	LAO 0020	1,000,000	92-93	Land Cover map of Laos
Air Quality Measurements				
Air Quality Observation	**********			
Demography				
Demographic Unit	**********			
Climate Zonation				
Agro-climatic Zone	LAO 0014	2,000,000	1968	Climatic Zones map of Laos
Administrative Bdys.				
Administrative Unit	LAO 0017	2,000,000		Provincial Map of Laos
Management Areas	**********			
Topography				
Elevation				MRC DTM's ?
	LAO 0004	5 minute		Derived from ETOPO5
Water Boundary / Body	LAO 0016	1,000,000	1988	River, Lakes, Islands map of Laos
	LM-DN 046	250,000	1955	Drainage Map - Si Mao
	LM-DN 049	250,000	1955	Drainage Map - Luang Nam Tha
	LM-DN 050	250,000	1954	Drainage Map - Lai Chau
	LM-DN 051	250,000	1954	Drainage Map - Dien Bien Phu
	LM-DN 052	250,000	1954	Drainage Map - Muong Ngoi
	LM-DN 013	250,000	1954	Drainage Map - Ben Giang
	LM-DN 017	250,000	1962	Drainage Map - Muong May
	LM-DN 029	250,000	1955	Drainage Map - Xaignabouri
	LM-DN 031	250,000	1955	Drainage Map - Muong Nan
	LM-DN 034	250,000	1954	Drainage Map - Luang Prabang
	LM-DN 035	250,000	1954	Drainage Map - Cua Rao
	LM-DN 036	250,000	1962	Drainage Map - Vang Vieng
	LM-DN 037	250,000	1962	Drainage Map - Khamkeut
	LM-DN 038	250,000	1954	Drainage Map - Vinh
	LM-DN 040	250,000	1955	Drainage Map - Thakek
	LM-DN 041			Drainage Map - Ban Don
Land Use				
Land Use Unit	LAO 0019	1,000,000	88-872-39	Landuse map of Laos

Appendix – 4: Lao Available Datasets

Geology				
Geological Unit	LM-GL 100	1,000,000	1988	Geology of Cambodia, Laos, Vietnam
Major Harvesting Activities				
Agriculture	**********			
Forestry	**********			
Mining Location	**********			
Fisheries	**********			
Water Quality Meas.	**********			
Soil Analysis Samples	**********			

Appendix – 4: Lao Available Datasets (cont.)

United States National Map Accuracy Standards

1. **Horizontal accuracy.** For maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch. These limits of accuracy shall apply in all cases to positions of well-defined points only. Well-defined points are those that are easily visible or recoverable on the ground, such as the following: monuments or markers, such as bench marks, property boundary monuments; intersections of roads, railroads, etc.; corners of large buildings or structures (or center points of small buildings); etc. In general what is well defined will be determined by what is plottable on the scale of the map within 1/100 inch. Thus while the intersection of two road or property lines meeting at right angles would come within a sensible interpretation, identification of the intersection of such lines meeting at an acute angle would obviously not be practicable within 1/100 inch. Similarly, features not identifiable upon the ground within close limits are not to be considered as test points within the limits quoted, even though their positions may be scaled closely upon the map. In this class would come timber lines, soil boundaries, etc.
2. **Vertical accuracy,** as applied to contour maps on all publication scales, shall be such that not more than 10 percent of the elevations tested shall be in error more than one-half the contour interval. In checking elevations taken from the map, the apparent vertical error may be decreased by assuming a horizontal displacement within the permissible horizontal error for a map of that scale.
3. **The accuracy of any map may be tested** by comparing the positions of points whose locations or elevations are shown upon it with corresponding positions as determined by surveys of a higher accuracy. Tests shall be made by the producing agency, which shall also determine which of its maps are to be tested, and the extent of the testing.
4. **Published maps meeting these accuracy requirements** shall note this fact on their legends, as follows: "This map complies with National Map accuracy Standards."
5. **Published maps whose errors exceed those aforestated** shall omit from their legends all mention of standard accuracy.
6. **When a published map is a considerable enlargement** of a map drawing (manuscript) or of a published map, that fact shall be stated in the legend. For example, "This map is an enlargement of a 1:20,000-scale map drawing," or "This map is an enlargement of a 1:24,000-scale published map."
7. **To facilitate ready interchange and use of basic information for map construction** among all Federal mapmaking agencies, manuscript maps and published maps, wherever economically feasible and consistent with the uses to which the map is to be put, shall conform to latitude and longitude boundaries, being 15 minutes of latitude and longitude, or 7.5 minutes, or 3-3/4 minutes in size.