Assembling the Building Blocks of Southeast Asia GIS Databases

Lawrence W. Crissman
Director, Spatial Data Projects
Griffith Asia Pacific Research Institute (GAPRI)
Griffith University, Brisbane, Australia

ACASIAN to APSDP, GAPRI

[Show HP]

The situation ten or twelve years ago

[Show ACASIAN SEA ADM Data]

East Timor added in detail later. Also Vietnam's new 1997 provinces.

[Turn on East Timor]

The situation now is that generally good data has been produced by many governments and NGOs, although accessibility varies greatly.

The following demonstration will present some readily available spatial datasets covering Southeast Asia (and sometimes the world) and the individual countries within it. The datasets that have been assembled for this demonstration are by no means exhaustive of those available from various sources, particularly at the level of particular countries.

Digital Chart of the World (DCW)
The DCW has been available for nearly ten years. It was originally released in the US Government's Vector Product Format (VPF) on a set of four copyright-free CD-ROMs costing US$200 that included the software necessary to load and display VPF and to query the data in a rudimentary fashion. Originally intended by the US Defence Mapping Agency (DMA) to be distributed free of charge, that low, low price was established in order to placate the British, Canadian and Australian government mapping agencies which are stakeholders in the included map information for some parts of the world in their neighbourhoods. By the mid-1990s, various GIS software firms had produced their own versions compatible with their proprietary software formats: ESRI (Environmental Science Research Institute), who had produced the original DCW for the DMA, claimed copyright on their ARC/INFO version that was distributed free of charge to customers, whereas Intergraph left their free MGE version in the public domain. Certain MapInfo resellers offered the data in that format, but at significant prices. More recently ESRI has made their version freely downloadable. Although their copyright technically remains in force, it would not be worth any effort on their part to enforce as it could always be claimed that anyone redistributing DCW data had converted them from versions that are in the public domain.

The DCW data are not altogether useful nor particularly easy to use, especially for medium to large areas of the world, because of their derivation and because of the way they are organised. The data in the DCW were all produced, on an automated basis, from the DMA's Operational Navigation Charts, a set of 1:1m map sheets that cover the entire world apart from Antarctica. The ONCs are designed for navigation by long-range, high-flying aircraft, and therefore represent features that have utility in that context. However, the smaller scale base map information they were generalised from are ultimately derived from the US Army Map Service work done during and after the Second World War, and therefore some features, coastlines in particular, are badly outdated in places like South Korea where extensive land reclamation has taken place. Presumably the aeronautical information they display is kept more current, but things like transport and hydrological features are not. Human settlements that are significant features when seen from high altitudes are mostly included, as are things like power lines and other 'cultural landmarks' such as docks, etc., that can be distinguished

from 35.000 feet. Yes, DCW contour lines are in thousands of feet (with some supplementary lines at 250, 500, and 750 feet), as that is the common unit for aviation use. Additionally, the DCW data are not completely uniform. For instance, the road and hydrological information on the ONC sheets covering the southern third of China is much less dense than comparable information on the sheets for North China, and those sorts of differences carry over into the DCW. These sorts of problems may have been resolved in what is essentially a new version of the DCW, the VMAP0 1:1m world datasets, but they are less readily and freely available than the DCW, and I have not yet worked with them.

The other basic problem with the DCW data, when used at a country-level at least, is that they are tedious to extract from the 5 degree square tiles that they are divided into in order to assemble those needed for particular countries or regions. There are around two thousand of such tiles altogether, and those that are needed for places the size of China or Southeast Asia number in the several dozens. The China in Time and Space (CITAS) project, which was active in the mid-1990s, extracted and combined the data covering all of China, which are still available from CIESIN (Consortium for International Earth Science Network), the CITAS sponsor, and from the China Data Centre, University of Michigan. My Australian Centre of the Asian Spatial Information and Analysis Network, (ACASIAN, now absorbed into GAPRI), did the same for the Republics of the Former Soviet Union and the Asia Pacific region, including Southeast Asia, partially alleviating this particular problem for certain areas of interest to ECAI.

Despite their deficiencies, the DCW data are still very useful for some ECAI purposes, primarily because they are the only readily available set of reasonably detailed world-wide base map spatial data that include many kinds of features, including coastlines, international boundaries, hydrography, hypsography, transport (roads, rails, seaports and airports), and cities and towns, plus other things of much less potential interest to ECAI.

[Show ponet for Southeast Asia, zoomed to the ug33 region, and then briefly turn on/off things like hydrography, roads, and settlements and built-up areas, with the following commentary]

The hydrography data are in fact too detailed for display at the level of even relatively small countries, and the main streams of larger rivers are not identified making it difficult to readily pick out and follow their courses. Roads are similarly quite dense, but are not usefully differentiated by grade or type. Settlements too are undifferentiated by size or status, although the larger ones have their built-up areas represented by polygons, however outdated they may be.

[Turn off all ug33 data, and zoom out to show ponet data for SE Asia.]

The 'ponet' (political-ocean) DCW data in particular, despite their somewhat antiquated and even in some places incomplete/erroneous status, are still extremely useful. For instance, the set of international boundaries can be used as a universally accessible standard, and indeed ACASIAN employs them for all datasets such as those for Southeast Asia which are intended to be juxtaposed with those for other countries. However, the DCW international boundaries in some instances do not agree with the territorial claims of certain states, such as those of China with respect to India in particular where the DCW represents the McMahon Line and other Indian claims. So, alternative boundaries representing the PR China version (or the rather different ROC version) of China's borders need to be substituted in some contexts.

Because they were produced from 1:1,000,000 scale maps with reasonably conventional resolution for that scale, the DCW coastlines for larger countries or regions are in fact too detailed to be appropriate for displaying on a computer screen or printing out on an A4 page. At the same time, they are rather too generalised for the analysis of small local areas.

[Zoom into the Hong Kong area to show some details]

The East View World Coastline Datasets

1:250,000 vector datasets with potentially at least 16 times the resolution of the DCW covering the whole world is being produced by NIMA (the US National Imagery and Mapping Agency) and distributed by East View Cartographic as the VMAP1 series. Although only around one quarter of the 250 or so tiles for all the other data are completed, East View have recently made world coastlines available for US$195 at a range of scales (resolutions, actually) from the new 250K VMAP1 up to 120M that go some way towards providing solutions to the problems of 1:1m resolution coast lines.

[Turn on the 250K coast for the Singapore region]

Note the far greater detail represented in the 250K data, including many more small islands.

[Turn off the DCW, and zoom out significantly]

However, when zoomed out enough to show even just the South East Coast of China, the greater detail is overwhelming and is in fact disadvantageous. Clearly, a lower level of resolution would be far more appropriate.

[Turn off the 250K data, and turn on the 3M, 40M, and 120M, consecutively, zooming more as you proceed from one to another, then zoom back to HK and turn them all on, including the 1:1M set, each with different colours.]

Esridata for the World

It is clear that the lower resolution datasets were produced from the 250K data by removing increasingly more vertices, the most crude method for generalising digital cartography and one that produces very jagged features yet does not adequately solve the problems of over representation of details. Far more sophisticated methods are available in some software programs that approximate the results of vectorising from smaller scale, less detailed maps. For instance, the data distributed with ESRI GIS products contains world data that I believe is the same as their old ArcWorld data which was at 1:3M scale/resolution..

[Turn on the Esridata for the World and turn off all but the 1:4M coastline data.]

Note the more rounded outlines AND the overall simplification of the details in the generalised ESRI low resolution data. ACASIAN has produced Low Resolution Datasets for China from 1:4m source maps that are similar, and which are designed for displaying and printing out maps covering all of China.

[Show the ACASIAN Low Rez China coastline from the Clearinghouse. Then zoom out and turn on the World Esridata world river data.]

The World Esridata river data show the major rivers only, including only the Mekong in Southeast Asia, and therefore may not be complete enough for some purposes. ACASIAN has produced low resolution river data for mainland Southeast Asia as well as China that may be more useful. Additional ones for Indonesia may be produced in the near future. (They also need to be downloaded from the Clearinghouse.)

[*Show the low rez SE Asian rivers along with the World Esridata Southeast Asian countries.]*

The gtopo2 DEM data

Digital Elevation Model (DEM) data are a very useful kind of world-wide dataset that are freely downloadable from the USGS (United States Geological Survey) and other www sites. However, they are raster (often called 'grid' data), not vector GIS datasets, and TimeMap View will not load and display them. Nor will ArcView or MapInfo without reasonably expensive extensions, such as Spatial Analyst or Vertical Mapper, respectively. Nor are the data formats in which the downloadable DEM are supplied compatible with the grid data formats of any ESRI GIS products, including ArcGIS 8, without processing procedures that would be beyond novice GIS users' capabilities. The gtopo2 DEM in particular required

manipulation of the file header format and entries that had to be figured out in the absence of any instructions whatsoever. ACASIAN now have all of those data, as well as the gtopo30 DEM data from tiles covering the world, discussed below, in formats that will load and display in ArcView 3x equipped with Spatial Analyst or in ArcGIS 8 (or ARC/INFO 7 Grid) and that can be converted for use with Vertical Mapper, and they will be provided gratis to ECAI participants on request.

[Display the gtotp2 DEM for SE Asia.]

The gtopo2 data contain altitude and ocean depth values in meters on a world-wide two arc minute grid, which equates to roughly three kilometers depending on latitude. Although I have recently found bathymetric datasets on an approximate 1 kilometer grid, I have not yet figured out what needs to be done to get them to load up in ESRI products. However, as you can see, the gtopo2 data have reasonable resolution when looking at regions the size of Southeast Asia. The ocean depth ranges have been set and the legend colours have been designed to display the critical sea levels before and after the three ocean surges associated with the end of the last ice age. The first major rise occurred at approximately 14,000 BP when the sea level rose from somewhere between 150 and 100 meters below today's level to 50 meters below, the second came around 11,000 BP when the oceans rose to 25 meters below present levels, and the third at 8,000 BP when the final surge pushed sea level to 5 meters above present levels. It then took 3,000 years for the oceans to gradually subside to approximately the current sea level. The whiter areas of the ocean represent depths less than 150 meters and then 100, 50, 25 and 10 meter depths. The darkest, almost black green is land 5 meters and less above present sea level, which was inundated from approximately 8,000 to 5,000 BP. Southeast Asia's Sunda Shelf now has around half the land area it had at the maximum of the last glacial period 18,000 years ago, and the coast of what is now China was also much further to the south and east of its present position, in North China as well as the southern portion shown in these DEM data.

[Zoom into someplace like Hainan Island to show the pixelation of the gtopo2 data.]

Etopo30 DEM Data

The approximately three kilometer grid of the gtopo2 DEM data make them pixelate when zoomed in to even fairly large places like Hainan or Luzon Islands. The etopo30 data are on a 30 arc second grid, which equates to somewhat less than one kilometer in mid latitudes, and therefore have grid cell sizes that are sixteen times smaller and provide correspondingly finer detail than the etopo2 data.

[Replace gtopo2 DEM with etopo30 DEM for the zoomed in area. Then zoom out to max and add the Southeast Asian DCW international boundaries and low res rivers.]

This is the finest resolution that was available in readily accessable, freely downloadable DEM data, although DEM for some places at much lower grid sizes has been ailable, at least to some. I have just at this conference been told that NIMA has now made Satallite Radar Terrain Modelling (SRTM) 100 meter grid data available for download in a user-friendly format. However, except for research on specific sites, such as sacred mountains, the thirty arc second etopo30 data are adequate for most ECAI purposes, and I eagerly await the day when TimeMap View will be equipped to handle raster GIS grid datasets such as the ones just demonstrated so that the altitude levels displayed can be freely manipulated by users. However, for now, users of TM View will have to be satisfied with static, un-manipulable raster images produced by screen dumps from DEM data displayed in other GIS software environments and then geo-referenced as well as possible. ACASIAN has some for China, and will produce some more for the ECAI Southeast Asia Team as a stop gap for the time being until TimeMap adds that capability.

NCGIS UNEP/GRID Administrative-Population Data

Back in the early-mid 1990s, I suggested to persons associated with CIESIN that, using the data ACASIAN was producing for Eurasia, I could produce grided population data for Asia on a one degree (or even much less) cell size, and that such a database if completed for the world would provide climate modellers and other earth scientists with a big advance over the 10 degree population data they were then using. Whether or not my idea had anything to do with the CIESIN decision soon thereafter to sponsor the National Center for GIS at the University of California, Santa Barbara, to hire someone to produce administrative unit and population data for Asia, I do not know. The resulting datasets the NCGIS produced also carry the imprinteur of the Consultative Group for International Agricultural Research (CGIAR), and were implemented by UNEP/GRID Arendal (UN Environmental Program/Global Resource Information Database in Norway), and are freely downloadable from UNEP/GRID web sites such as www://grid.no/cgiar/htmls/asiademo.html. They cover all of Western, Central South, East, and Southeast Asian countries, as well as all African countries, those places being the regions of the world for which such data were not already available to the scientific community at the time.

[Display the NCGIS spatial data for Southeast Asia, as well as one attribute table, briefly.]

On the face of it, these data look really good, and would seem to be of great use to ECAI. However, note that the size of the administrative units vary greatly from country to country. Note also from the attribute tables that the base year for the population figures also varies, and that the figures for the latest year are all estimates produced by uniformly augmenting the base year figures by some percentage. Presumably, the administrative data should be current to the date of the base population figures, although I don't think that is actually the case. In any event, this is not documented so far as I know. Nor is the source of the boundary information or the procedures used to match them up with the coastline and international boundaries, which are all from the DCW. However, some inferences can be made about the latter from a close examination of the NCGIS data, which appear to have been rubber sheeted to the coastlines and international boundaries at mostly appropriate places but not to any internal features. This is clear from the case of Timor, where the point where western border of the exclave of xxxxx has been incorrectly placed at the western border of the main portion of East Timor, squishing the latter up into the eastern part of the island.

[Zoom into the island of Timor on the NCGIS data, and then superimpose the ACASIAN data for East Timor)

This procedure also produced some large anomalies in the planimetric accuracy of the NCGIS data for large countries like China, which are off by dozens of kilometers when looking at data at some distance from a coast or international boundary. Typical spatial distortions in the NCGIS data are displayed on the ACASIAN www pages at www.asian.gu.edu.au/compare.html, which also discusses some major errors and anomalies in the county-level population figures for China. To make a long story short, I do not trust the accuracy of either the GIS data or the associated population figures for the Asian data produced by the NCGIS and distributed by UNEP/GRID. The African data, which comes from a different source, may be quite OK. The similar data produced by ACASIAN for Southeast Asia are, I believe, more trustworthy even though some are not as accurate as could be desired due to errors in the maps used in their compilation.

[Turn on the ACASIAN SE Asia admin datasets]

Apart from the Philippines data produced at James Cook University in Townsville, Australia, which are not complete for Mindanao and have not been regionalised at higher administrative levels, the ACASIAN produced administrative datasets for Southeast Asia often exist at multiple administrative levels and usually include capital/seat points. For instance, for Myanmar (Burma), the ADM2 level is displayed as it has units that are approximately equivalent in size to the ADM1 level in other Mainland Southeast Asian countries. There are only 10 ADM1 divisions and states in Burma, which are far more extensive than the provinces of Thailand.

[Turn on the Burma ADM1 on top of the ADM2 regions.]

The same CIA map was used by ACASIAN and NCGIS to produce our respective administrative data for Myanmar, but it contains only the top level divisions and states and the very numerous townships, with nothing in between where there should be an ADM2 level.

[Turn on the ADM3 on top of the ADM1 data, and then add the ADM3 seat points.]

Collaboration with academics from Myanmar who had access to government data on population allowed ACASIAN to construct the regions for the intermediate ADM2 district administrative divisions and to include population estimates for a number of years.

[Turn off the ADM1 and ADM3 Burma data, leaving the ADM2 districts, and then display its attribute table.]

In this case, the ADM2 Myanmar data are the appropriate ones to use when displaying and analysing all of Mainland SE Asia, as they are more or less equivalent in size to the ADM1 units in the other neighbouring countries.

The data displayed for Malaysia were produced at the University of Western Australia, and include AMD2 distrik and ADM3 mukim regions (where the latter exist).

[Briefly turn on the Malaysia ADM2 and ADM3 data.]

ACASIAN/GAPRI has not produced its own data for Indonesia, apart from some preliminary work on the 1970s vintage administrative system. What is displayed here are the NCGIS data regionalised to produce the ADM1 provinces and with problems like those on Timor fixed up.

Recently Griffith University and the Badan Pusat Statistic (BPS, or Central Statistical Body) in Jakarta have formalised an agreement that allows GAPRI to redistribute BPS spatial and statistical data for all of Indonesia. As received from BPS, the spatial data had various kinds of errors, mainly with regard to the placement and size of some islands and island groups, which took a considerable amount of time to identify and correct using the VMAP1 coastal data as a guide in order not to miss any small equivalent islands.

[Turn on the original BPS data in conjunction with the DCW ponet data in a different color, zoomed first to the region NW of Kalimantan and then to the West Coast of Sulawesi.]

These BPS data include the ADM1 propinsi, ADM2 kabupaten/kodya, ADM3 kecamatan, and ADM4 desa, there being approximately 70,000 of the latter. However, these data are not inexpensive, as the total price to the BPS for a complete set is 23 million rupiah.

[Display the corrected ADM2, ADM3, and after zooming to Java, the ADM4 data.]

Somewhat similarly, there are administrative data for Cambodia produced locally under UN sponsorship that can be obtained from government sources, but for only US$100. They go down to ADM3 level, and also have capital/seat points.

[Pan to Cambodia, and sequentually turn on ADM2 and then ADM2 caps, ADM3 and ADM3 seats.]

The Cambodian data also include railroads, main roads, minor roads, and hydrography.

[Show in turn, as above.]

The latter are roughly as dense as the DCW hydrography digitised from the 1:1M ONCs, but of course the two datasets only align approximately.

[Turn on the ug33 hydrography for comparison.]

There are known to be equivalent locally produced datasets for other Southeast Asian countries that also include lower administrative levels, but they have not yet been obtained by GAPRI. The ECAI Southeast Asia Team may be able to source such datasets for Laos, Thailand, and Vietnam.

In summary, it can now be said that fairly accurate and reliable GIS spatial data are readily available for most if not all countries of Southeast Asia. The days when one had to be prepared to vectorise one's own spatial data, except for very local places where great detail is needed, are well and truly in the past. Obtaining access to data that is known to exist can still be problematic for some countries in the region, but ECAI is increasingly in a position to facilitate spatial data access if it cannot in fact provide quality data through the Metadata Clearinghouse.