Review of establishment of 50 K Topographic Vector Database

By

S.K. Wijayasinghe

Senior Deputy Surveyor General (Training)

Former Superintendent of Surveys (GIS) from 2001-2006

Abstract

In Sri Lanka, the Survey Department entered the world of digital mapping and GIS technology, by converting the topographic map series into digital form and establishing a digital topographic vector databases. The department has completed establishing 1:250,000 scale database in 1999 and 1:50,000 scale database in 2003, and establishment of 1:10,000 database is in progress.

Creating and building a sound Topographic Vector Database for a country is a huge challenge faced by any National Survey and Mapping organization. This article describes how the Sri Lanka Survey Department established Topographic Vector Database on 1:50,000 scale (called 50K database) in a record time span of 2 ½ years from mid 2001 to end of 2003. The Survey Department obtained an award for special achievements in GIS at the ESRI International User Conference held in Sandiego, California in 2004.

Unlike a printed map in which all the data are in one paper layer, the topographic vector databases are designed to comprise many different data layers. In case of 1:50,000 database, there are 9 layers, namely Building, Transport, Hydrography, Places, Terrain, Administration, Land Use, Reserves & Grid. Another layer called 'Control' was added to accommodate geodetic control points established by GPS.

To achieve this task, the Survey Department of Sri Lanka, has successfully developed and implemented a methodology, with minimum manual interaction, to digitize the existing topographic maps on 1:50,000 scale, covering 65,610 sq.km. to achieve a highest degree of accuracy possible.

This methodology was developed using scanned images of maps & combined positives, geo-referencing them in CAD environment, screen digitizing them in CAD, converting the CAD drawings into data exchange format and converting them into ArcInfo coverage format, using AML (Arc Macro Language) programs.

This was a task which required the highest level of technological applications as well as the highest level of management techniques. The technology itself had to be managed and the management itself needed technology.

The ultimate objective was to change from the era of manual production of maps into digital production of maps form the topographic databases and for GIS applications.

1.0 Introduction

The Survey Department of Sri Lanka which was established in the year 1800 by the British colonial administration, which is also the national mapping organization, has established a geodetic control network and mapped the whole country during the first half of the 19th century and it has legal mandate and the responsibility for producing and supplying topographic data for the users.

The most popular topographic map series was the 1” map series on the scale of 1 inch to 1 mile. It contains all the topographical information and geographical names that are in practical use at the time of preparation of the map series in respect of natural or man made topographical features, place/village names, administrative areas etc. There are 72 maps in this series.

After Sri Lanka converted itself in to the metric system of measurements in 1982, a new topographic map series was prepared on the scale of 1: 50,000. This was basically a metric version of former 1” series with up to date information collected from Aerial Photographs and/or from field. There are 92 map sheets in this series. The first edition of this series was completed in mid 1990s.

The Survey Department is also in the process of preparation another series of topographic maps on a larger scale of 1: 10,000.

With the advent and development of computers, information technology, data storage capacities and database management systems, a method was emerged in the world to store map information in a computer. This created a new era in mapping called ‘Digital Mapping’ since the information in a map is stored in a computer in digital form.

The department has commenced digital data compilation by photogrammetric methods in 1992, and commenced digitizing topographic maps in the late 1990s. The department has completed digitizing 1:250,000 scale map and 1:50,000 topographic map series, and 1:10,000 series is now in progress. With these data, the Survey Department has established digital topographic vector databases for GIS applications which can be supplied for the users.

The maps that were exclusively printed documents drawn on a flat sheet of paper, depicting objects in the real world before, is now converted into digital maps. Once the spatial information in a map is combined with descriptive information and stored in a computer database, such systems are called Geographic Information Systems (GIS) and the database is called a Topographic Vector Database.

Creating and building a sound Topographic Vector Database for a country is a huge challenge faced by any national Survey and Mapping organization, and the Survey Department faced this challenge successfully by establishment of Topographic Vector Database in the form of ArcInfo coverages.

The topographic vector databases comprise many different data layers that varies slightly depending on the scale.

The databases are in the form of ArcInfo coverages and organized in the form of tiles as shown in the grid index for topographic mapping. The data can be supplied either as separate tiles or even combination of many tiles and in the form of original data format, or as shape files or as DXF files or export (e00) files.

2.0 The Designing of Databases

The primary and most important activity in establishment of a Topographic Vector Database is to design the database structure. The data model that was selected was ESRI ArcInfo coverages, which was proven to be the technically most popular model at that point of time.

An ESRI ArcInfo Coverage is a georelational data model that stores vector data; i.e., both the spatial (location) and the attribute (descriptive) data for geographic features. ArcInfo_Coverages use a set of feature classes to represent geographic features. Each feature class stores a set of points, lines (arcs), polygons, or annotation (text). Feature attributes are stored in the ArcInfo_Coverage's .adf files. Other attributes can be stored in INFO tables or tables in an RDBMS, then joined to features with a layer or a relationship class. ArcInfo_Coverages can have topology, which determines the relationships between features.

An ArcInfo_Coverage is stored as a directory. The directory name is the ArcInfo_Coverage name. An organized collection of ArcInfo_Coverages is called a workspace. An ArcInfo_Coverage stores a set of thematically associated data considered to be a unit. A single ArcInfo_Coverage usually represents a single layer, such as soils, streams, roads, or land use. In an ArcInfo_Coverage, features are stored as both primary features (points, arcs, polygons) and secondary features (tics, links, annotation). Feature attributes are described and stored independently in feature attribute tables. Each ArcInfo_Coverage workspace has an INFO database stored under a subdirectory, named info. Each .adf file in an ArcInfo_Coverage folder is related to a pair of the .dat and .nit files in the INFO folder. The arc.dir file in the INFO directory is used to keep track of which pair of .nit and .dat files is related to which .adf file.

More than one feature class is often required to define the features in an ArcInfo_Coverage. For example, line and polygon feature classes both exist in an ArcInfo_Coverage representing polygon features. Polygon features also have label points, which appear as a separate feature class. Every ArcInfo_Coverage has a feature class containing tic points, which represent known, real-world coordinates. These tic points help define the extent of an ArcInfo_Coverage; they do not represent any actual data points within the ArcInfo_Coverage.

The set of features contained in an ArcInfo_Coverage depends on the type of geographic phenomena being modeled. The types of feature classes that can be found in an ArcInfo_Coverage include:

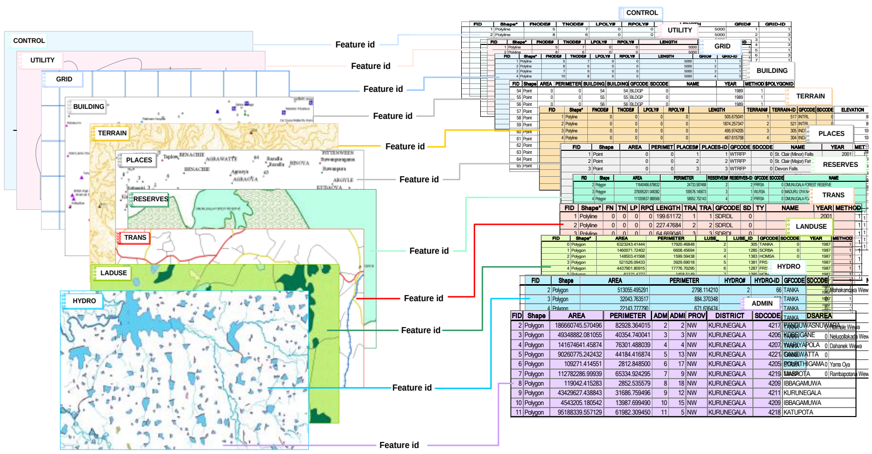
- *Point* -- A point defined by an x,y coordinate pair used to represent point features or assign User-IDs to polygons. Used to establish point locations such as well sites, and mountain peaks.
- *Arc* -- A line defined as a set of ordered x,y coordinates used to represent linear features and polygon boundaries such as street sections, contours, streams, sewers, power lines, and gas lines.
- *Polygon* -- An area defined by the arcs that make up its boundary, including arcs defining any islands inside. User-IDs are assigned to label points. Polygons represent area features such as soil units, land use, parcels, building footprints, forest stands, and ownership. Attributes for a polygon feature are found in an attribute table named pat.adf.

2.1 Design of 50K Database

This 50K database is tile structured, and each tile contains 9 layers, namely BUILDING, TRANSPORT, HYDROGRAPHY, PLACES, TERRAIN, ADMINISTRATION, LANDUSE, RESERVES & GRID. Additional layer called CONTROL was introduced to accommodate Geodetic Control points and the data was obtained from a different source.

LAYER NAME	DESCRIPTION	FEATURE TYPE
BUILDING	Important Buildings with names	POINT
TRANSPORT	All Transport Features including Main Roads, Secondary Roads Tracks, Foot Paths, Railway Lines etc.	ARC
HYDROGRAPHY	All water features including Streams, Tanks, Reservoirs, Lagoons, Sea, Internal & external Islands etc.	ARC & POLY
PLACES	All point features including Village Names, GN Division Names, Place Names, km posts etc.	POINT
TERRAIN	All contours & spot heights at the contour interval of 20 m or 100 ft	POINT & ARC
ADMIN	Administrative limits covering Province, District & DS Divisions	ARC & POLY
LANDUSE	Landuse features showing major landuse catogories	ARC & POLY
RESERVES	Reserved areas covering Forest & Wild Life reserves	POLY
GRID	Grid lines	ARC

The design of the database is described in the Data Dictionary which also gives specifications of the database, coding, feature definitions etc. Data in each layer is linked with an attribute table having items such as Geographic Feature Code (GFCODE), Survey Department Code (SDCODE), NAME, TYPE, YEAR, METHOD, and ELEVATION (for terrain only)



<u>Database Structure</u>

- **GFCODE (all coverages)**
- **SDCODE (all coverages)**
- **Name (only in some coverages)**
- **Type (only in Transport & Utility coverages)**
- **Year**
- **Method**
- **Elevation (only in Terrain coverage)**

The time of data collection of original data in this database may vary from 1983 to 2002 which varies from tile to tile.

2.2 Feature Coding

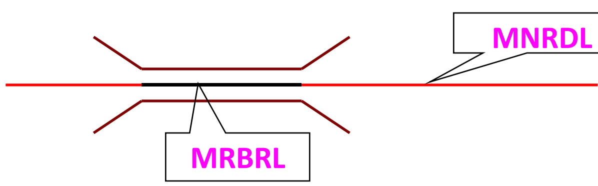
Selection of feature coding is an important aspect in designing the databases, as each topographical feature has to be coded in order to identify them uniquely and unambiguously.

GFCODE

Unique in the Database (5 Alphabetic Characters)

- **1-4 Chars : represent the feature**
- **5th Char: represent the feature type**
  - P -Point
  - L - Line
  - A - Area
  - V – Virtual

Eg. MNRDL – Main Roads
MRBRL – Main Road on Bridge



Feature Codes in Transport Layer

FEATURE	COVERAGE	RC GFCOD
Main Roads (LINE)	TRANS	MNRDL
Main road on Bridge (LINE)	TRANS	MRBRL
Main road on Bund (LINE)	TRANS	MRBNL
Main road along Tunnel (LINE)	TRANS	MRTNL
Main road on Causeway (LINE)	TRANS	MRCWL
Secondary/Minor Roads (LINE)	TRANS	SDRDL
Secondary/Minor Road on Bridge (LN)	TRANS	SRBRL
Secondary/Minor road on Bund (LINE)	TRANS	SRBNL
Secondary/Minor road along Tunnel (LN)	TRANS	SRTNL
Secondary/Minor road on Causeway(LN)	TRANS	SRCWL
Secondary/Minor road on Dam (LINE)	TRANS	SRDML

Further, there are different feature types such as points, lines and polygons that may be given same feature code. For example, the geographical feature 'TANK" has to be represented by line feature as well as polygon feature, both of which should be coded as 'TANK". Therefore the unique feature code was designed to have 5 alphabetic characters 4 of which are to represent the feature itself and one to represent feature type.

TANKL to represent the outline of a tank
TANKA to represent polygon of the tank

This logic can be applied to even to show Road features as polygons in creating larger scale databases like 1: 1000 and also to represent wider Expressways on even small scale.

The type code 'V' was used to represent virtual lines, which are imaginary boundaries used to break polygons when necessary. For example, when a stream falls into a tank the linear feature between stream and tank should be a virtual line as there is no physical boundary in between.

3.0 Technical approach for Digitizing

A printed map of 1: 50,000 scale contains six colours, namely, Black, Red, Blue, Green, Yellow and Brown to represent different topographical features.

Black color represents, all transport features including Main Roads(casing), Minor Roads (casing), Rail tracks, Tracks, Footpaths, all Place Names, km posts, Bridges, Culverts, Land use boundaries and Buildings.

Red color represents, Main Roads (Fill), Administrative Boundaries, and Administrative Division Names.

Blue color represents, all water features and their names.

Brown color represents terrain features including contours and heights.

Green and Yellow represented filling of various landuse areas which do not represent any feature boundaries.

There are combined positives for each of these colors, which were used to make printing plates for offset printing of the maps. Although a printed map is on a piece of paper which is liable to expand or shrink with the passage of time, these positives are on transparent and very stable material.

Therefore the highest accuracy in digitizing can be obtained by digitizing those combined positives and not by digitizing paper maps. Further advantage of using positives for digitizing was separation of topographical features. For example, when the digitizing was done on black positive the operator was sure that he does not digitize water features or contours. One of the most difficulties in digitizing maps is identification of topographical features correctly. If identification is incorrect, wrong feature can go to the wrong layer, with incorrect code.

Those combined positives were scanned and geo referenced by a method called rubber sheeting in CAD.

Digitizing was done on CAD environment as screen digitizing and the separation of features was done while digitizing itself by putting different features on different layers of CAD.

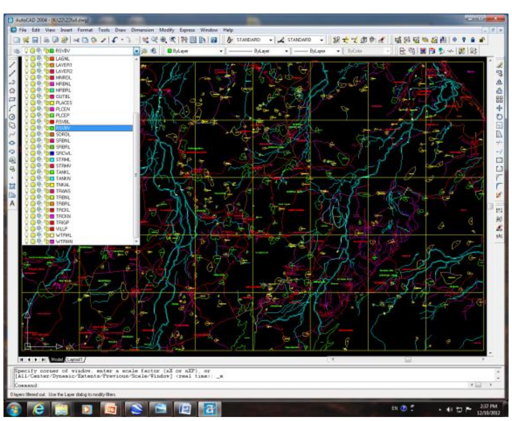
The layers were named by the feature code to be given in the database. For example the feature code for Main Roads is MNRDL and all the Main Roads were digitized in a CAD layer called MNRDL, and once the CAD is converted in to coverage all the features in MNRDL layer gets feature code MNRDL.

Since the contours are mostly continuous except on the places where the contour value is printed, the digitizing was done using a software that can convert raster to vector automatically or semi automatically, and it was much faster than ordinary screen digitizing.

In case of contours, and heights digitizing was done in 3D environment, where, Index contours in INDXL layer, Intermediate Contours in INTRL layer, Spot heights in SPHTP layer and the elevation (height) of contour or point was in Z value of the contour or point so that feature code and the elevation is directly derived from the layer name and the Z value with the AML.

Different features are placed in different layers and are assigned with different colors and line types very similar to the way those are shown in maps. This helps in identifying and distinguishing features quickly and easily.

Those CAD drawings were converted into Data Exchange Format (DXF) in ASCII form with 16 decimal places of accuracy and then converted in to Arc Info Coverage format using AMLs written exclusively for that purpose.



4.0 Management Approach

The digitizing was done by 10 operators, who represented different combinations in accuracy and speed, some were very fast and very accurate, some were fast but not very accurate, some were slow but very accurate, and some were slow as well as not very accurate.

The number of operators had to be limited to 10 primarily due to lack of resources in terms of computer hardware and software.

Digitizing is generally laborious, time consuming and boring work, but with the introduction of technical approaches and management approaches, the boring work made interesting.

The progress of digitizing was monitored in terms of number of km digitized in case of boundaries, and the number of points digitized in case of Place Names and Heights. In general, progress of women were much better than men in digitizing, and the men were in the category of slow but accurate.

Digitizing and creating the database was done in a stage wise approach. The topographical features that has demand from the users and that are easy to handle were taken up first. In the first round the digitizing was done for Transport, Hydro, Place Names, and Administrative Boundaries and Buildings using the Black, Red and Blue positives, and in the second round contours were digitized using the Brown positive, and finally the landuse using Black positive again. In the Landuse coverage, certain boundary features in Transport and Hydro coverages are repeated to form landuse boundaries.

Most of the time in the ½ year in 2001 was spent on designing the database, testing technical approach and methodology, testing AMLs, testing accuracy in geo-referencing, and digitizing, and basically for the planning of the project.

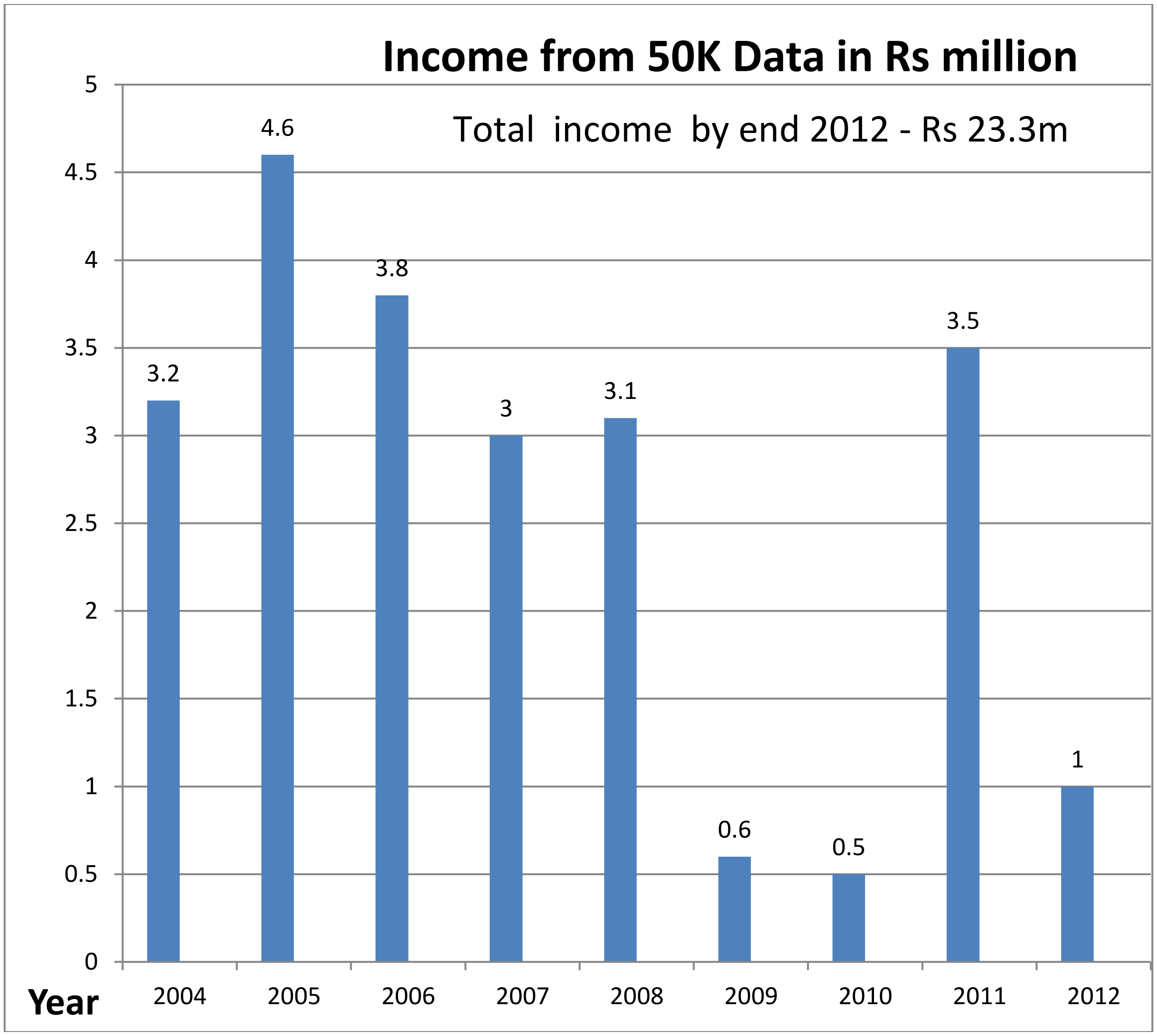
Possibility of errors in digitizing was of 3 folds. The first being the possibility placing the digitized feature in an incorrect layer, which was minimized by digitizing one type of feature at a time. For example, when Main Roads in a map sheet were digitized, the other features were not digitized until Main Roads were completed.

The second being connectivity errors with either gaps or overlaps which are not easily seen in ordinary CAD environments. However, a customized program helped to identify such connectivity errors in the CAD environment itself.

The third being the accuracy of digitizing itself, or in other words the number of turning points digitized in a boundary feature.

5.0 Services rendered by 50K database

A simple way of proving the services provided by the 50K digital data is by analyzing the user demand for the data. The following chart indicates the amount of sales since 2004.



Even at the very nominal rates given for data, there is an income of Rs 23.3m by the end of 2012, which clearly indicates the user demand. The actual demand is much more and intangible, in view of free issue of data for various State Organizations. The mapping activities in relief and rehabilitation operations after boxing day Tsunami in 2004 was based on 50K database, and it was of immense use at the Centre for National Operations (CNO) established under the direct supervision of Her Excellency the President at the Presidential Secretariat, Colombo.

5.0 Conclusion and lessons learned

At the end, this project involved digitizing 99,325 km of linear transport features, 96,253 kms of linear Hydrographic features with 20,269 number of Hydrographic areas, 118411 kms of landuse boundaries, with 90,505 number of different landuse areas, 282,349 km of contours, 3171 height points 13,927 km of administrative boundaries, 525 administrative areas, 215,359 number of buildings, 58,332 number of place names, km posts, culverts and other names.

This was a task which required the highest level of technological applications as well as the highest level of management techniques. The technology itself had to be managed and the management itself needed technology.

Possible human errors in digitizing were identified in advance during the planning stage and attempts were made and approach was designed to minimize such human errors.

Better planning can achieve better results. The time consumed for planning is not a waste as long as it can produce better results.

A good team work approach was another factor for the better achievements.

Digitizing and creating the database was done in a stage wise approach instead of handling all the data layers together, which made the life easier, simpler and minimized possible errors.

Maximum use of facilities in technology and required customizing can expedite the output drastically, while making the operators feel more comfortable and convenient in their laborious work.

Any boring and laborious work can be made interesting by introduction of technical approaches and management approaches appropriately.

In the end, this task followed by establishment of 10K database, created an environment that can change from the era of manual production of maps into digital production of maps form the topographic databases. In other words what is updated in the future is the topographic database and not the topographic maps, and the updating of maps should be done from the updated database.

References: ESRI Web site