Database generalization and production of derived maps at 1:100000 and 1:250000 scales using NTDB in NGII context

Durgendra M. Kayastha

Abstract: *Multi resolution geodatabase will support diverse application requirements ranging in scale and resolution. Creation of such a database primarily enhances production of derived maps, which is one of the key application areas within NGII context. The paper identifies the processes involved specifically in the area of production of derived maps.*

Introduction

National Geographic Information Infrastructure (NGII) may be defined as the technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and non-profit sectors, and the academic community.

Building of geospatial information infrastructure reduces duplication of effort among agencies, improve quality and reduce costs related to geographic information. The objectives of such an infrastructure is to make geographic data more accessible to the public, to increase the benefits of using available data, and to establish key partnerships among data producer and users to increase data availability.

Geospatial data, at least the framework data (basic topographic data) forms the basis upon which other geospatial data may be built. In order to support diverse application needs at varying level even the basic framework data need to be made available at several resolution or scales.

In the context of NGIIP (National Geographic Information Infrastructure Project), production of lower resolution databases has been planned in addition to NTDB supporting production of smaller scale topographic maps as well as regional and national level planning activities. Preparation of database and subsequent production of maps at 1: 100000 and 1:250000 are briefly outlined in this paper.

Base data (NTDB)

Basic framework data is organized at sheet level and are maintained as different layers. The principal data classes are Topography, Building, Transportation, Landcover, Hydrography, Utilities, Administration, Designated Areas, and Vilname.

The NTDB is based on the basic topographic maps that are of two scales. Data based on 1:25000 sheet cover 7’30” by 7’30” area while that based on 1:50000 sheet cover 15’ by 15’ area. The difference in data content is according to the change in scale.

Within each data class the data layers are actually maintained as follows:

Class	Coverage			Remarks
	Point	Line	Area	
Topography	Topog_pt	Topog_ln		Contours and spot elevation
Transportation		Trans_ln	Trans_ar	Roads and trails
Landcover			Landc_ar	Landcover types
Hydrography		Hydro_ln	Hydro_ar	Streams and rivers
Designated Area		Desig_ln	Desig_ar	National parks and reserves
Utilities		Utili_ln	Utili_pt	Electricity Lines
Administration		Admin_ln	Admin_ar	Administrative boundary
Ward	Ward_pt	Ward_ln	Ward_ar	Ward boundary
Place Name	Vilname			Place names
Building	Build_pt		Build_ar	Isolated building and built-up areas

Specification for 1:100000 and 1:250000 mapping

Topographical Survey Branch, Survey Department has already prepared mapping specifications for 1:100000. The specifications provide for the data types that the database will contain. However there will be some changes in the specifications to adapt to the requirements in terms of the database.

There also will be some departure from the specification previously prepared by the Survey Department in cases of symbols. The symbol specification will be prepared conforming to the possibilities in the software used.

Data Model

Data model for NTDB100 and NTDB250

Items marked * will not be implemented in NTDB250

Class: Transportation
- Subclass: Road
  - Subsubclass: Highways, Feeder roads, District roads, Other roads,
- Subclass: Major trail (only in the mountains)
- Subclass: Railway
- Subclass: Ropeway
- Subclass: Airport
  - Subsubclass: Runway *

Class: Building and Settlement
- Subclass: Building *
- Subsubclass: Religious building
- Subsubclass: Others *
- Subclass: City/Village

Class: Topography
- Subclass: Contour
- Subclass: Characteristic point
  - Subsubclass: Peak, Pass
  - Subclass: Spot elevation

Class: Landcover
- Subclass: Cultivation *
- Subclass: Vegetation
  - Subsubclass: Forest
  - Subsubclass: Bush *
- Subclass: Other
  - Subsubclass: Snow

Class: Hydrography
- Subclass: River
- Subclass: Glacier
- Subclass: Lake
- Subclass: Canal
- Subclass: Structure *

Class: Place Names

Class: Administrative Area
- Subclass: Country
  - Subsubclass: Region
    - Subsubsubclass: Zone
      - Subsubsusubclass: District
        - Subsubsubsubsubclass: VDC/Municipality

Class: Designated area
- Subclass: National Parks/Wild life reserves

The contour data will be maintained at an interval of 100 metre with index contour at 500 metres in all cases irrespective of the area being covered contrary to 50 metres for the plains as specified in the specification for 1:100000 maps. Similarly the contour interval will be 200 metres in the case of 1:250000 with index contour at 1000 metres.

Naming Convention

Following the sheet numbering system, data pertaining to a particular sheet will be extracted from the NTDB and stored in a new folder with the parent directory name as NTDB100. The sub folder name will be the sheet number itself slightly rearranged. That is the alphabet part of the sheet number and the number part itself is switched in order to represent the folder name. The naming convention for the coverage in NTDB100 will be as follows (e.g. \\NTDB100\A2787\to100_ln).

Data Class	Coverage Name			Feature Type
	Point	Line	Area	
Topography	To100_pt	To100_ln		Contours and Spot elevation
Building	Bu100_pt	Bu100_ln	Bu100_ar	Settlement
Land Cover	La100_pt	La100_ln	La100_ar	Landcover area
Hydrography	Hy100_pt	Hy100_ln	Hy100_ar	River edges
Transportation	Tr100_pt	Tr100_ln	Tr100_ar	Roads
Designated Area	De100_pt	De100_ln	De100_ar	National Park and protected areas
Admin Area	Ad100_pt	Ad100_ln	Ad100_ar	Regional/Zonal/District/VDC/Ward
Place Name	Vi100_pt			Place names

Similarly, data layers pertaining to a particular 1:250000 sheet will be extracted from the NTDB100 and stored in a new folder with NTDB250 as the parent directory name followed by the first four number of the sheet number preceded by the alphabet part of the sheet number as sub-folder name. The naming convention for the coverage in NTDB250 will be as follows (e.g. \\NTDB250\E2781\to250_ln).

Data Class	Coverage Name			Feature Type
	Point	Line	Area	
Topography	To250_pt	To250_ln		Contours and Spot elevation
Building	Bu250_pt	Bu250_ln	Bu250_ar	Settlement
Land Cover	La250_pt	La250_ln	La250_ar	Landcover area
Hydrography	Hy250_pt	Hy250_ln	Hy250_ar	River edges
Transportation	Tr250_pt	Tr250_ln	Tr250_ar	Roads
Designated Area	De250_pt	De250_ln	De250_ar	National Park and protected areas
Admin Area	Ad250_pt	Ad250_ln	Ad250_ar	Regional/Zonal/District/VDC/Ward
Place Name	Vi250_pt			Place names

Utilities and controls data will not be used at this database level.

Basic Steps for 1:100000 and 1:250000 database and mapping

Following is the general steps to be followed for 1:100000 and 1:250000 database:

1. Generalization by feature class
2. Editing of geometry and attributes including matching of features horizontally and vertically
3. Final database preparation
4. Map design
5. Map composition
6. Printing of Maps

Generalization by feature class

Stepped generalization procedure may be followed that means generalization will be made in several steps. The first step would be to generalize NTDB feature to NTDB100 features. Second step is the generalization of features of NTDB100 to NTDB250 features. Similarly the lower resolution level database will be obtained by generalizing features set of the immediate higher resolution database.

Generalization entails two fundamental steps viz. selection or deletion of features and geometric processing for thinning, displacement etc.

Based on the feature class/subclass encoding in the NTDB database, several features may be deleted, which will not be represented in the NTDB100. But this general approach is only a part of the process, as this process might not eliminate or might remove certain features that may need to be represented in the NTDB100. Hence a rigorous approach of encoding additional attribute to individual vector may be performed which enables selective and controlled extraction of required data only.

Once required data at higher geometric resolution but at reduced content is obtained the remaining work is to process for obtaining right geometry and topology. The process includes geometric thinning of vertices and editing for topology and intra-layer matching.

Editing of geometry and attributes including matching of features horizontally and vertically.

With the generalization of vectors and related area and point features, the shift in the position of vectors will result in mismatch of features. Similarly small area feature may have to be merged with the adjacent large area feature. Small line segments of different attribute may require to be changed, e.g. feature (line segment) depicting a bridge may be removed or attribute or feature code changed to match with contiguous line feature etc.

After these processes, one more run of edge matching will have to be performed to match feature between the sheets. All these changes have to be made interactively more or less to match features and attributes vertically and horizontally.

Map database preparation

The process entails draft printing of maps at desired scale and checking for consistency of data as to density, requirement of important features and names at the right places etc. and correcting the discrepancy and finally archiving the dataset as NTDB100 dataset. This is used for further application i.e. in our case production of 1:100000 maps.

Similarly, based on NTDB100 database, NTDB250 database will be generated following the same procedure. NTDB250 database will be used for further application such as production of 1:250000 maps.

Map design

Designing of map layout (i.e. preparation of style sheets including symbols, text, diagrams etc.) will be carried out taking into consideration the mapping specification for 1:100000/1:250000. Specification for 1:100000 previously prepared by the Survey Department and the software and hardware that are available will also be taken into consideration, in addition to the desired output that could be obtained in the given situation

Several programs and symbol design will have to be done at this phase.

Map composition

Suitable routine will be prepared to automate the production process minimizing user intervention to the limit as far as possible. One should however bear in mind that total automation might not be feasible at this point.

Printing of Maps

This is the final stage of the 1:100000/1:250000 mapping where the output will be generated using the routines prepared and based on the database. Out put will be a hardcopy pull out using available plotter.

Generalization approach

Places

Places should be selected to represent the general pattern of settlement distribution as far as possible. Hence the names of places should first be selected based on importance of such places at different levels, such as national, regional, zonal, district, and VDC and Municipality levels.

In order to achieve this additional attribute field NML may be coded in each of the place name. The domain of values for NML may be obtained by using the following table.

Level of importance	Head-quarters	Religious	Commercial	Touristic	Historical	Others
National	10	11	12	13	14	15
Regional	20	21	22	23	24	25
Zonal	30	31	32	33	34	35
District	40	41	42	43	44	45
VDC/Municipality	50	51	52	53	54	55
Local	60	61	62	63	64	65

Based on the value of NML selective extraction of place names could be facilitated. Further it should also be considered that no two places should be closer than certain threshold distance. The value of threshold depends on the database resolution level. In case of NTDB100 the threshold distance may be 1000m.

Transportation network

Due to rugged nature of the terrain, different types of transportation routes have varying significance and are closely related with the area. For instance, in the northern mountain region where there is no access to vehicular road, several trails will be more important as these are the only means of movement. This is not the case in the southern plains. Besides, it is important that roads and trails should be shown to places depicted in the map.

It is therefore necessary to improvise certain methods to support identification of such important trails at different levels of importance. A straightforward approach would be to introduce additional attribute field showing importance level.

Buildings and Settlements

Religious buildings, factories, etc should be coded with additional attribute value showing their importance. Other buildings may be processed to obtain distribution pattern. Depending on the cluster size point or polygon could be created.

Topography

Depending on the specification, contours are selectively extracted. Further processing is performed as to geometric generalization and editing for inconsistencies such as small ring contours, crossing of contour lines, mismatch with the spot elevation and other features etc.

Hydrography

Streams and rivers in the NTDB database should be coded with stream order value. This will provide basis for selective extraction of streams and rivers. The extracted vectors will then be processed for generalization of shapes as required.

Landcover and Hydrography

First step would be to combine some of the landcover subclasses. The resulting data will be examined for small areas. If necessary, such small areas will be merged with adjoining larger areas. In some cases, applying constrained growth, small areas of importance may be enlarged. Once a suitable generalization in areas is obtained, the vectors will be processed for line thinning and final data will be created. Similar approach will be applied for Designated Areas as well.

Administrative Area

Boundary segments will be generalized followed by the polygonization to obtain required area coverage.

Further Testing

Conceptually the methods have been identified. Some of the generalization can be done in raster environment while many others could be achieved within the vector-processing environment. Most of the process experimented upon could be programmed to minimize interactive sessions. The efficiency of the methods and the data models need further refinement.

Conclusion

As the testing process has not been finalized yet, a thorough result can not be explained yet. But the document has outlined the processes involved. In order for the testing work to progress a task group may be assigned with the necessary equipment and software for the purpose. Actual details on the share of automation and interactive intervention can be decided after all phases of testing have been completed. However, a 1:100000 scale and 1:250000 scale topographic layer on NGII topographic database have been planned. This will also give a tremendous support to derived topographic mapping of Survey Department at these scales.

References:

Esri, 2002. Working with Geodatabase: Powerful Multiuser Editing and Sophisticated Data Integrity, 9 p.
Esri, 2000. Map Generalization in GIS: Practical Solutions with Workstation Arcinfo Software, 19 p.
Jaakola, Olli, 1997. Quality and Automatic generalization of Land Cover Data. Academic dissertation, Department of Geography, Faculty of Science, University of Helsinki, 154 p.
Kilpelainen, Tiina, 1997. Multiple Representation and Generalization of Geo-Databases for Topographic Maps. Doctoral dissertation. Publication of Finnish Geodetic Institute No. 124, Helsinki, 229 p.