CONSTRUCTION PROGRAM OF DATABASE WITH VECTOR DATA FOR 1:25,000 SCALE GEOGRAPHICAL INFORMATION

Hiroyuki Ohno, Mamoru Koarai, Toshiyuki Terabayashi, Shinpei Ishigaki
Eiichi Tamura, Koji Otsuka, Kiyoaki Nakaminami
Geographical Survey Institute
1-Kitasato, Tsukuba-City, Ibaraki-pref, Japan
TEL: +81-298-64-5915 FAX: +81-298-64-3056
E-mail: dave@gsi.go.jp

Abstract

Largest scale of topographical map all over Japan provided by governmental organization, Geographical Survey Institute, is "1:25,000 scale Topographical Map". It has been managed and revised with raster type data format recorded each map sheet in a computer file. The management system of this topographical map will be changed with vector type data format recorded on seamless database. Features of this database are following:

(1) Elimination of dual management about same information
(2) Management with vector type data set for all with time series management
(3) Seamless database of the whole of Japan without neat line and acceptance of time management
(4) Acceptance of data record format that integrates geometry shapes and topology information
(5) Separating on the specification between the function of cartographic symbol drawing and abstracted data definition

Background and circumstances of construction of new management system with vector type geographical information database are described, and specification of the database and construction schedule is also described.

Introduction

The digital geographical information such fundamental spatial data has been managed separately from cartographic information drawn on papers. They have some decisive differences each other. One difference is the notion of data definition. The definition form of digital geographical information for fundamental spatial data, equivalent in used for GIS, is abstracted from real feature, such as centerline vector of roads, point data for control points and so on. The other topographical map information is cartographic symbol, for example, the roads drawn by two parallel lines and control points drawn by triangle symbol. Another difference is topology information. The former has topology information, but the latter doesn't always need to have it. Their two kinds of data have been compelled to manage separately by reason of that. If their information could be integrated to same data, it will be very rational to manage them, nevertheless it has not been integrated yet on national level data set in Japan.

The national largest scale map covered all over Japan is "1:25,000 Topographical Map", produced by Japanese governmental organization, Geographical Survey Institute (GSI). Some digital geographical information data sets have already been published, which were gotten from 1:25,000 Topographical Map, such as administration border line digital

data and many kinds of national digital information. They were not produced only by GSI, but also by National Land Agency at the time. However, almost national digital information has not been updated but a few data set which have small data quantity. The problem not to revise digital information suitably is caused by difficulty of revising digital information directly, so that all conversion from topographical map information to digital information had been taken as method to make new revision of digital information. Fig.1 shows the current general flow to revise for digital information. Production of the digital geographical information such fundamental spatial data becomes strongly important, and it needs to provide with great low costs or free as national infrastructure on computer network like the Internet, all over Japan. It has already been adopted a resolution by the Cabinet meeting as the policy, so that the GSI, our organization, must provide them with all rapidly.

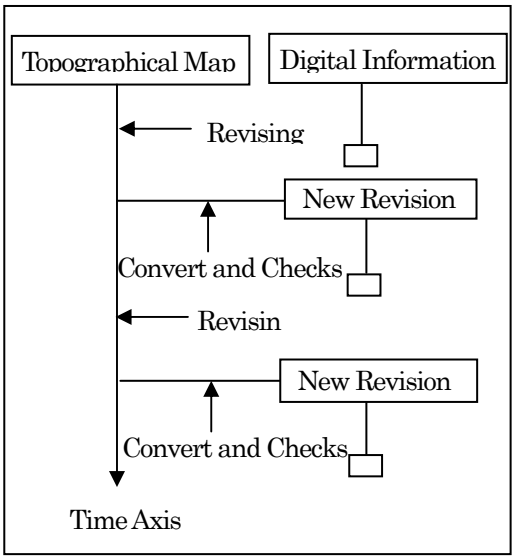

Fig 1. Current Revision Flow for Digital Information

At this moment, the former problem is eminently important and has strong efficiency. About 4,300 map sheets of 1:25,000 Topographical Map covers all area in Japan. There are many features in the real world which are moved, constructed, destroyed or changed, therefore their changing information have to be revised exactly on the data set. Consideration for all the possibilities of the methods to do their revision, but if all topographical map data are converted to digital geographical information data set every times of revision, it is very clear to be unreasonable plan. Effective and realistic solution for integrate both data revising flow has been needed with new notion include data handling, conversion, format and digital geographical information itself.

Currently, 1:25,000 Topographical Map data are managed by raster data format recorded on cartridge magnetic tape each map sheet. On the other hand, the specification of the fundamental spatial data generally takes vector type data format. It is impossible to integrate in their entirety. At the final analysis, transferring data types of either side to same types of the other completely, and constructing new geographical information database of 1:25,000 level with integrated data forms between fundamental spatial data and topographical map data have been decided as main component of New Topographical map Information revising System(NTIS).

Requirements and Circumstances

At first step, requirements and current circumstance were analyzed.

One of the strongest and most important requirement was elimination of dual management about same kind of data. It is feared unconformity would be caused, because their management is done independently at each working section such as lettering, control point information etc.. For example, in the case of lettering information, following investigation were carried out. There are two kinds of lettering information, lettering raster data for topographical map and lettering table for database which have already been provided for all each other. The former is simple bitmap data, and the latter is

defined as corner coordination list of rectangle outside along the lettering. The former can be generated from the latter on investigation so that it is decided to integrate into the latter and include only the latter in NTIS.

The other of the most important requirement was perfect cartographic drawing from abstracted vector data. It was quite difficult problem (Fig.2). Investigation what kind of attribute is required to draw has been being carried out in a bid to be perfect drawing. For example, the road centerline vector has its attribute value of width on the map, flag of tunnel, bridge, both side condition and so on.

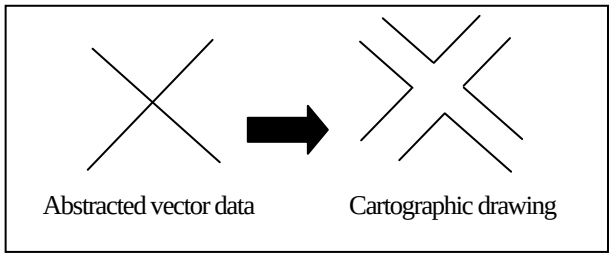

Fig.2 Cartographic drawing from abstracted vector data

More than 99.9% of these problems would be solved, but 0.1% problems left let to be difficult this problem. The current specification for cartographic drawing on the 1:25,000 Topographical Map was defined in 1986. However, 1:25,000 Topographical Map are being revised all by hands, so that no map sheet is drawn based on the specification perfectly. Though that is a few case on each sheets, it is estimated total of over 10,000 cases, and various kinds of exceptions on all 4,300 map sheets. It is impossible to research all exception case. If all exception cases were defined, it must not develop functions to solve there all cases by reason of development costs. A simple vector specification, which only used drawing to display like cartographic drawing in exception case, is added to data specification of NTIS, and limited functions are developed to accommodate main exception cases only.

Next, time series management has to be investigated. Time series management can let us to realize deference management between one time surface and the other time surface. Changing information on topographical map is very useful to decrease amount of data into map users to update their information, and it is acceptable to scientist and analyzer who use some old topographical map at past some time surfaces. Of course, past topographical map information can not manage on this database, but the data of future time surfaces can be referred freely. This may be going to change the method to provide revising information to map users, which provides difference information only with computer network like the Internet.

Third investigation, the Geographical Survey Institute has 10 local offices in Japan. Topographical map revising has been also carried out at their local offices, so that this database is estimated to be revised at not only main office but also local offices. On the other hand, the same database would like to be located at all local offices, because of avoid dual management. All 10 local offices will manage data of each managing area only, and main office at Tsukuba-City and one local office at Chiyoda-Ku, Tokyo, have whole data in Japan as backup records with function of replication of DBMS. This means three same data set will be managed for data backup with form of dispersion database as if it appears only one database from external. Replication has to be carried out automatically, and doesn't have to conscious of other location databases.

Fourth, it must be estimated that how many records and amount of recorded data will be managed. All information of 1:25,000 Topographical Map will be managed in this database. Some research was carried out. As a result of their investigation, there are about 50,000 to 100,000 features on topographical map sheets each. It will be increased by conversion to data record on this database under the specification (refer the "Basic Data Type and Policy of data

Specification" about details of specification), number of records will be perhaps 100 million order for all Japan and amount of data will be 100 GB order initially. It is estimated that this database will be used about 10 years in the future. It is decided that this database will manage 1 billion records and 1 TB data amount.

Fifth, circumstances around 1:25,000 Topographical Map and digital geographical information were analyzed. One basic definition around Japanese survey will be changed at same time when NTIS will finish constructing. This is conversion of geodetic datum, Tokyo Datum to Japanese Geodetic Datum 2000 (JGD), so that all coordinates must be transformed with parameters. This conversion let to move neat lines of all map sheet, and the topographical information managed by raster format don't have advantage to convert smoothly. If conversion with raster formed topographical information were enforced, complex transformation about each pixel data would be estimated total of over 9.6trillion pixels on all topographical maps, and no one would be changed current management system.

Last analysis, it is estimated to suit new technologies such as time management, direct digital survey with D-GPS, especially the computer networks. This may be revolution about digital geographical information include for GIS. Digital mapping means generally digitizing from various base maps almost always their sources are paper maps equivalent non-digital information, in other words non-digital information is the first and digital geographical information is the second. What is the revolution with new technology? Position of digital geographical information on the notion is reversed to the first. Directly getting the digital geographical information is possible to change the notion about survey. The direct digital geographical information should be defined as vector information.

As a result of all analysis, new geographical information database becomes to have solution of above requirements.

Concept

NTIS can have some subsystems as its component connected local area network (LAN). New geographical information database will be constructed as one of the NTIS subsystem. Construction of the new geographical information database started on third quarter in 2000, at the same time, prototype making with new vector type data for topographical map information was also carried out. Judging from the results so far obtained and analysis result of the requirements, this database should be constructed under 5 concepts as following:

(1) Elimination of dual management about same information.
This is common concept. All subsystems of NTIS are constructed under this concept.

(2) Management with vector type data set for all with time series management.
All managed data are defined as vector data under NTIS specification. Lettering, contour, center lines of road and railways, and other all features on 1:25,000 Topographical Map.

(3) Seamless database of the whole of Japan without neat line.
All data within this database are managed as if they are not cut off at crossing point with neat lines logically, so that it can be assumed seamless geographical information database.

(4) Acceptance of data record format that integrates geometry shapes and topology information.
NTIS has another subsystem to revise digital geographical information like CAD software. Topology information required by GIS don't always be needed when the digital geographical information are revised. Topology information can calculate from condition of geometric shapes and connection flag, so that this database has no topology information clearly. However, the condition of geometric shapes and connection flag has to get

perfectly. Topology and geometric shapes integrated style was accepted.

(5) Separating on the specification between the function of cartographic symbol drawing and abstracted data definition.

This database doesn't have drawing function, and provide only function of data management. For example, functions of calculation of topology information, cartographic symbol drawing, and revising information are defined as function of application side. This database has simplified data specification for easy revising of data set of all geographical information.

Basic Data Type and Policy of Data Specification

The data specification of this database aimed to be simplified format it could. As topology information is integrated to geometric shapes, there are only 2 basic data types in the specification. They are point and arc. Polygon will be calculated from condition of geometric shapes that circumscribed by closed arc and polygon control point inside the arcs. The coordinate value is defined as integer by 0.0001 seconds unit of latitude and longitude with JGD.

All data records have a common table and an additional table. The common table consists of 8 fields (Table 1). First three fields means record ID itself, fourth field has 3 time tables and means existing terms on topographical map, fifth field has 2 time tables and means effective terms on the database itself, seventh field is defined for speedy data search. Additional table is defined for each kinds of features. Point type data record are assumed the first two fields, Longitude Value and Latitude Value, defined as the position of coordinate itself, and arc type data record are assumed the first two fields are starting point, and ending point and array of middle points have to be defined on additional table. Time Table has three time fields, generation time, elapse time, and confirmation time when this feature is confirmed to be in real world. Database Time Table consisted of fields of creation time and deleted time on this database. The former is used to generate deference data during specified term, and the latter is used to manage by database itself.

Longitude Value
Latitude Value
Duplication ID
Time Table
Database Time Table
Record ID (Reserved)
Rock Index
Quality Value

Table 1, Common Table for each records

Three types of trus-shape polygon is defined by combination of point and arcs. First type is the simple polygon drawn by only one arc defined right side is surface. This type is defined for large number features such as cartographic symbol

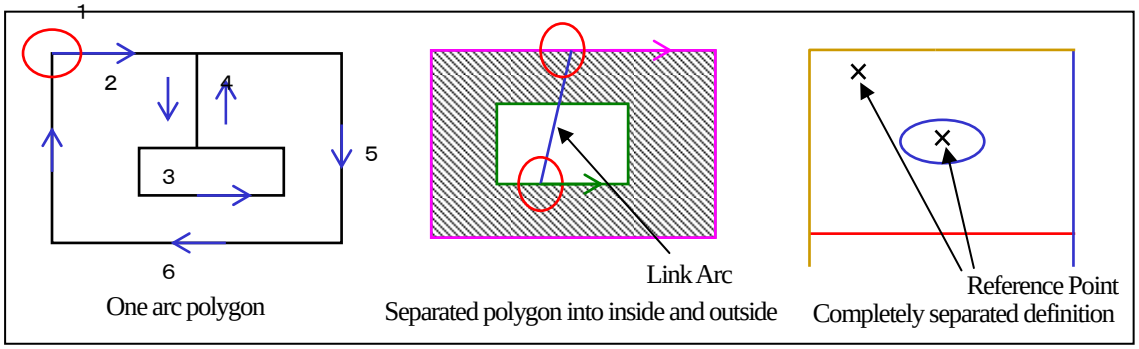

Fig.3 Three definition type of polygons

of small building. This is main specification of polygon records on this database for the reason that it is very easy to get outline vector with conversion from raster to vector. Second type is separated polygon into inside and outside. Both side polygon are defined as same as the first type polygon, and they are connected by “Link Arc”. This type is used by polygons painted hatching inside such as cartographic symbol of large building. The first type torus-shape polygon can’t be drawn hatching well. Third type consists of completely separated arc and reference point of each polygon. This type must calculate to define each polygon at application side. This type is only used for the definition of autonomous community, because they are not defined by simple border arc generally.

Attribute of each cartographic symbol is recorded in “Additional Table” following the Common Table of each data record shown as Table 2. First field consists of some attribute fields defined each kind of cartographic symbols, records information of necessary value for drawing, for example, road arc has road width as cartographic symbol on topographical map, attribute of tunnel, and so on. Second field references external table. This external table always manages conditions of the real world, such as national road numbers, names of autonomous communities, maybe includes various value which features on the real world will have in the future. Third field is recorded table id for reserved table. This will be used in the future.

Attribute Values
Attribute Index to External Table
Attribute Index to Reserved Table
…

Table 2, Additional Table following the Common Table

Searching Method and Lock System

Database design was also carried out under investigation to speed up function to search and lock. All data on this database have at least one information of coordinate. It is anticipated to search with area which defined as list of coordinate, so that range partition with coordinate, longitude and latitude, was taken as the method to speed up searching function. All data were separated to manage by tables located each limited area, searching function can read only near requested area from database by referencing this range partition. However, if too many data records were managed with one range partition, the searching performance might be decreased. Number of data records that managed with one range partition have to be defined suitable number for capacity of systems, especially hardware storage.

There is one appropriate problem to revise with database by many workers at once. This was double revising at the same area. It is possible to cause a discordance of recorded information. Therefore suitable lock system must be defined under management of database. A polygon indicate rocked area was defined as one kind of information on this database. Database management system can refer this polygon information to know who is rocking in which data area. The other user has already rocked requested area from another user to rock, database management system can denied access. Here, suitable rock system was defined, too long line records and too wide area records are possible to cause another problem. Especially, counter line is always defined as very long data record, for example, counter lines of 100m high may be over 1,000km around all over Japan. If such this record was rocked by one user, the other user who want to revise same counter at faraway area, this area almost always don’t have relation with the area where rocked by first user. Therefore, all line records were defined to cut off by short length. This length will be decided by investigation of the hardware performance on about July 2001.

How about index for rock to be judged whether one record should be rocked? There is the fields for Rock Index at seventh field on common table of each records, already written at “Basic Data Type and Policy of Data Specification”

section. This is defined with coordinate value as follows:

Point record: The coordinate of itself is also defined as coordinate value of Rock Index.

Arc record: If this record have no complement point, coordinate of starting point is defined as coordinate value of Rock Index. If this record have odd complement points, coordinate of middle complement point is defined as coordinate value of Rock Index. If this record have even complement points, coordinate of one before middle point is defined as coordinate value of Rock Index. Rocking system would refer only this Rock Index when rocking request is received.

Implementation of Drawing Method of Cartographic Symbol

This is not the specification of this database, however, the NTIS is constructed to revise the 1:25,000 Topographic Map, so it must be guaranteed to draw cartographic symbol on Topographic Map from the vector data recorded on this database directly. In 2000, a test to draw from their data under specification of this database was carried out with about suburb area of covered by 90 map sheets on Ooita prefecture and urban area of covered by 4 map sheets at Nagoya City.

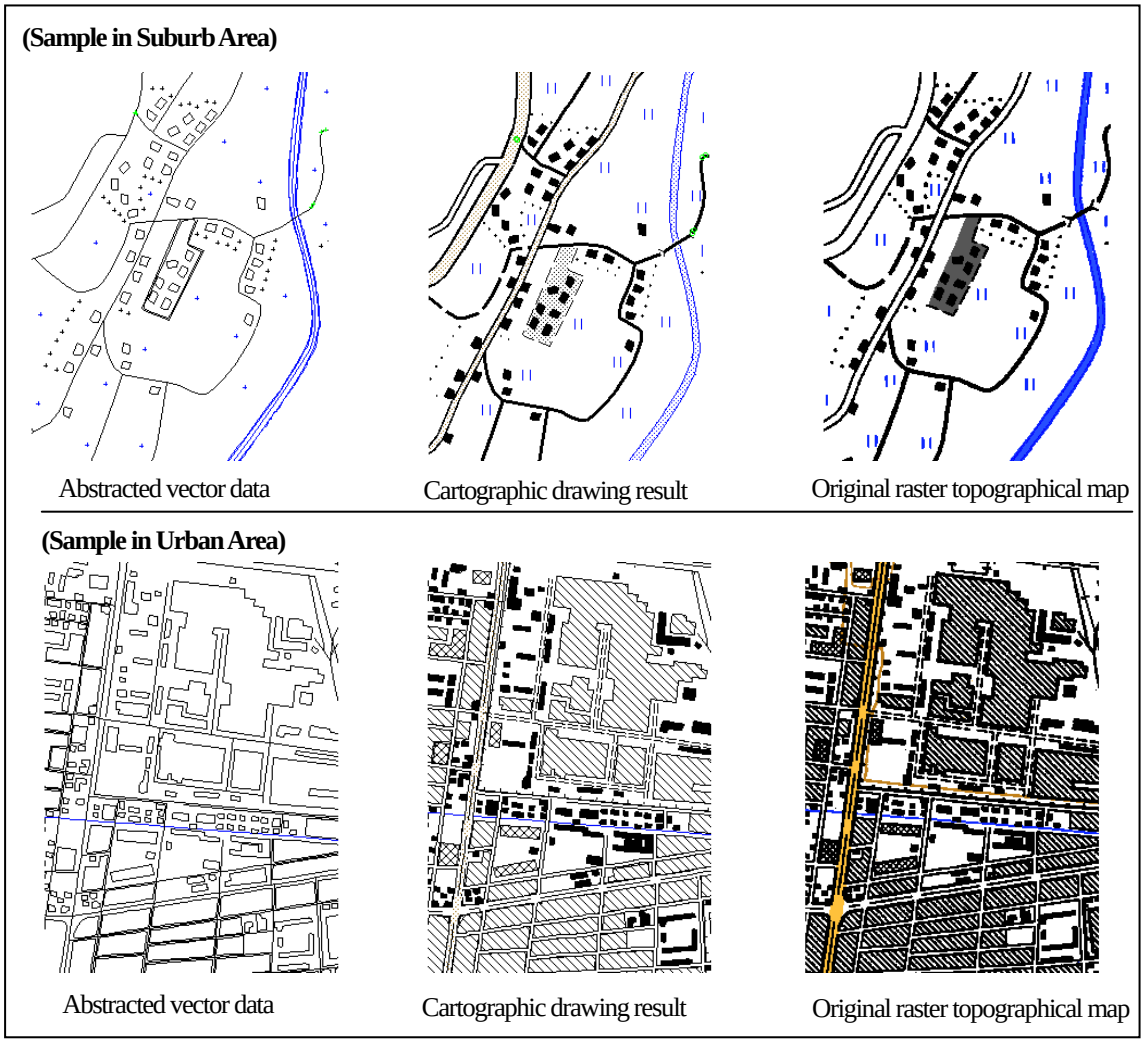

Fig.4 Result of test drawing for investigation of abstracted vector specification

Especially, drawing of road crossing point is the most important point. There were many drawing problems. But, number of their problems have been being decrease day by day according to improvement of specification of abstracted vector data record format and development functions to avoid problems. The case of exception is drawn by simple vector line.

There is one problem left about cartographic drawing of road crossing. The case of overhead crossing, especially more than three-level interchange, how to draw suitable cartographic symbol. A attribute "Priority of drawing" is defined for road and railway vector records. This attribute means which record should be drawn upper than the others. This is not absolute value. This is defined relative value at each crossing point. Application should draw according to this "priority of drawing" value. Another cartographic symbol can be drawn by each drawing process with implementation of function of cartographic drawing on application. This can draw almost a carbon copy of current raster cartographic information. This means vector data can use as original data to draw cartographic symbol.

Time Schedule

The specification of this database has been decided. Data conversion from raster data to abstracted vector data has already started since April 2001. Area of about 3,300 map sheets, about 75% area covered in Japan, will be converted to abstracted vector data in this year, and remain will be converted in 2002. At same time, application, revising subsystem of NTIS, which has functions to draw cartographic symbol and to revise their information are constructing now, and it will be finished constructing on March 2002. Management system of Japanese topographical map will be transferred completely in 2002.

Conclusion

Management system of largest scale map covered all area of Japan, 1:25,000 Topographical Map, will be transferred to new management system with vector type data and database. This perhaps means revolution of notion to revise all geographical information includes cartographic field and GIS field, and must change methods to provide geographical information.

Current specification of this database have anything more than attribute for cartographic drawing, but it will start a discussion about new specification of topographical map drawing suit cartographic drawing from abstracted vector data. We think that information of additional attributes for to define about various features existing in real world will add to current specification.

Reference:

1. Geographical Survey Institute, 2000, Investigation Report for ideal 1:25,000 Scale Vector Geographic Information under National Standards for Geographic Information
2. Geographical Survey Institute, 1986, Cartographic Symbol Specification for 1:25,000 Topographical Map
3. Geographical Survey Institute, 2000, Working Report for Making Vector 1:25,000 Topographical Map Information in Oita Area
4. Geographical Survey Institute, 2000, Investigation Report for Updating with Deference Information for GIS Fundamental Information