

RESEARCH ON THEMATIC DATA PROCESSING AND DATABASE CONSTRUCTION OF SEQUENCE DIAGRAM

Tingting ZHAO ^{1,2}, Xiuli ZHU ^{1,2}, Wanzeng LIU ^{1,2}, Wei MA ³, Yunlu PENG ^{1,2}, Ye ZHANG ^{1,2}, Xinpeng WANG ^{1,2}, Yuewu WAN ^{1*}, Xinli DI ¹

¹National Geomatics Center of China, No.28, Lianhuachi West Road, Haidian District, Beijing 100830, China - (zhaotingting, zhuxiuli, lwz, wyw, dixinli, pengyunlu, zhangye, wangxinpeng)@ngcc.cn

² Key Laboratory of Spatio-temporal Information and Intelligent Services (LSIS), MNR, No.28, Lianhuachi West Road, Haidian District, Beijing 100830, China

³ National Quality Inspection and Testing Center for Surveying and Mapping Products, No.28, Lianhuachi West Road, Haidian District, Beijing 100830, China - maweichn@qq.com

Commission IV, WG IV/7

KEY WORDS: Thematic Data, Data Processing, Database Construction, Quality Requirement, Sequence Diagram

ABSTRACT:

The sequence diagram in the atlas shall represent the latest macro overview achievements of the country, closely focusing on the major national development strategies, and select topics directly related to economic construction and people's life, including thematic information such as sea area, railway, highway, population, tourism, forest, nature reserve, surface coverage, urban agglomeration and strategic layout. According to the selected thematic content, the relevant thematic information is collected and analyzed, and the available thematic information data is extracted through data normalization and spatial processing, and then spatially matched with the thematic geographic base map to form a thematic map database. This paper studies how to deal with the thematic data in the sequence diagram and how to build the database, including thematic data analysis, thematic data processing, thematic database construction and quality requirements.

1. INTRODUCTION

The sequential map in the atlas should represent the latest macro-level results of the country, with content closely focused on major national development strategies and selected topics directly related to economic construction and people's lives, including thematic information on sea areas, railways, roads, population, tourism, forests, nature reserves, land cover, urban clusters and strategic layout.

This paper examines how the thematic data in the sequential map is processed and the database is constructed, including thematic information analysis, thematic data processing, thematic map database construction and quality requirements.

2. THEMATIC INFORMATION ANALYSIS

The public authoritative topic information comes from various related departments, which is restricted by the sharing and exchange system, so it is difficult to obtain all the topic information. This paper mainly uses the data information publicly released by professional departments and the information collected by other projects as the main source of information. Due to the problems of various types of information, irregular data and non-uniform indicators, the workload of data fusion, screening and extraction, and determination of indicators is large, and information from

thematically relevant materials should be selected for analysis and processing before use.

The analysis of the information led to the following main sources of information being used for the sequence map thematic data.

1) Results of the country monitoring project

As a major national survey, the Geographic National Condition Census is an important means to obtain a comprehensive picture of the current situation and spatial distribution of China's natural and human geographic elements, and is a fundamental exercise to grasp the basic situation of surface nature, ecology and human activities. The geographic national survey was completed in 2015, with the completion of the data on the results of the first national geographic national survey and the data on the results of the basic monitoring carried out annually since 2016. The thematic information collected and processed by the project can be used as a reference data source for the thematic information of the sequential map.

2) Various professional materials

Specialist information refers to specialist geographic information data published by authoritative departments, which is used for relevant element change discovery and data collection to further improve the current status of data results. It can be used to verify the quantity of thematic information in the sequence map and other attribute information to ensure the correctness of the thematic database. Specialist information mainly includes.

* Corresponding author

- Ministry of Water Resources' water census results, river and lake reservoir code information.
- Ministry of Transport's road data, road names and coding information, railway line and station names and coding information.
- Notification of administrative division changes, place name changes, results of administrative division boundary surveys, administrative division names and administrative division codes from the Ministry of Civil Affairs and civil affairs authorities at all levels.
- The Ministry of Natural Resources' list of national geoparks, botanical gardens (arboretums), nature reserves, forestry sites (forest parks), wetland parks, and the list of national key ecological function areas.
- For airport data from the Civil Aviation Administration, please also refer to the published list of public airports.
- Data on scenic spots from the Ministry of Housing and Urban-Rural Development.
- Sixth Census information from the National Bureau of Statistics.
- The Ministry of Culture and Tourism's list of the 1-7 batches of National Key Cultural Heritage Protection Units for information resources on topics such as cultural heritage monuments and revolutionary memorial sites.

3. THEMATIC MAP DATABASE CONSTRUCTION

In order to reflect China's valuable natural and cultural resources, a series of thematic information on nature reserves, world cultural and natural heritage, and historical and cultural cities are selected to reflect the harmonious development of man and nature.

The thematic map database, based on thematic geographic base maps, is represented in categories and layers according to thematic categories and geometric features of elements, and contains roughly 20 thematic categories and more than 40 data layers.

3.1 Thematic Map Database Design

3.1.1 Thematic Map Database Structure and Content Design: The spatialisation of thematic information has to be based on basic geographic information data as a spatial skeleton and positioning basis, so before producing thematic maps, the main element layers in the general geographic map database are first used to produce thematic geographic base maps.

On the basis of the thematic geographic base map, the coordinate system of the thematic map database is determined, and the data stratification and naming, attribute table structure, data storage format and other contents are determined with reference to the national common geographic map data building requirements, so that the thematic map data can be standardized and systematized, which is convenient for the management and later maintenance and update of the thematic map database.

The selection of topics in the thematic database include World Regions, World Terrain, China Regions, China Terrain, China Images, China Seas, "One Belt, One Road", China Railways, China Highways, China Airports and Ports, China Cities, China Population, China Ethnic Groups, China Climate, China Water

Systems, China Vegetation, China Geological Hazards, China Natural Sites (China National Nature Reserve, China National Forest Park, China National Scenic Spot, China National Geological Park), China Cultural Heritage (World Cultural and Natural Heritage, China Tangible Cultural Heritage, China Key Cultural Relics Protection Units, China Intangible Cultural Heritage) and China Cultural Sites (China National Historical and Cultural Cities, China National Historical and Cultural Villages, China National Historical and Cultural Towns, China Chinese Revolutionary Monuments). (WANG et al., 2018b; ZHAO et al., 2020)

3.1.2 Thematic Data Layer and Attribute Table Structure:

The thematic database includes 20 selected element classes and 47 data layers. The thematic data layers are named with four characters, the first character "Z" represents the thematic data, the second and third characters are the abbreviation of the data content in Hanyu Pinyin, and the fourth character represents the data model type. The geometric type of the vector data is point P, line L and surface A. The raster data type is R. (WANG et al., 2018a)

According to the content of the thematic map expression and the characteristics of the thematic element attributes, the structure of the attribute table is designed and the attribute items are defined, etc., which are summarised in detail in Table 1.

Topic	Layer	Thematic Content	Attribute
World Region	ZSZA	National administrative divisions worldwide	Name Dcode Cpname
World Terrain	ZSDR	Undulating terrain around the world	
China Region	ZZQA	The provincial administrative divisions of China	Name Dcode PRname
China Terrain	ZDSP	The names, heights and publication dates of the major peaks in China.	Name Elev Dcode Mcode PubDate
	ZDSL	The names, ridge lengths, etc. of the 1st, 2nd, 3rd and 4th class mountain ranges in China.	Name MonLen Mname Elev Mcode Dcode
	ZDSA	Plains, plateaus, basins, etc. On a scale of 1 to 2, with grades divided by area.	Name TArea Dcode Ttype
	ZDSR	Colour layered landform halo renderings of the country.	
China Image	ZYXR	Indicates national 30m resolution Chinese image data.	

China Sea	ZHYP ZHYA	Territorial sea baseline points, island (reef) points, and four major sea locating points. Sea island (reef)	Snum Name Type Dcode Belsea Belisd Sarea Lat Lon
	ZHYR	Generated from ETOP1 data.	
One Belt, One Road	ZDLP	The important towns through which the Road passes.	Name Ccode
	ZDLL	Road routes	Name Type
China Railway	ZZTL	China's major railways	Name Rtype Rnum Rlen Beladm
China Highway	ZGLL	Divided by type into highways, national roads, provincial roads, etc.	Name Rnum Type Rlen
China Airport and Port	ZJTP	Indicates our major airports, ports, crossings, etc.	Name code Dcode Type Lev Ftype
	ZJTL	Nautical line	Name RLEN
China City	ZCSA ZCSP	Faceted cities Dotted cities	Name Class Crate Dcode Popu Care Lev
China population	ZRKA	Population with the county administrative unit as the mapping unit.	Name Popu Mpopu Wpopu Cpopu Rpopu Pdens
	ZRKP	Urban population linked to urban points, urban population of towns and cities above the county level.	Name Class Dcode Cpopu Popu
China Nationalities	ZMZP	Punctuated ethnic distribution.	Ethn Popu
	ZMZA	Facultative ethnic distribution.	Ethn Popu

China Climate	ZQHL	Isotherm, isoprecipitation.	Valu Ftype
China Water	ZSGP ZSGL ZSGA	Flood protection works. Agricultural water construction (irrigation districts). Water Transfer Project (South-North Water Transfer). Hydropower projects	Name Rname Hcode Descr Exted Parea Ftype
	ZHLP ZHLL	Point systems of water such as springs, hot springs, wells and reservoirs. Rivers, canals, drains, etc.	Name Dcode Lev Rcode Svol Ftype
	ZSHA	River reservoirs and lakes.	Name Lev Rlen Rcode Warea Svol Ftype
	ZLYA	There are 10 basins at level 1, 69 at level 2 and 195 at level 3 (Ministry of Water Resources)	Name Lev Bcode Belsea Inout
China Surface Coverage	ZDBA	Using GlobeLand30 data, 10 categories of land cover elements are represented: arable land, forest, grassland, shrubland, wetland, water bodies, tundra, artificial ground, bare ground, glaciers and permanent snow	Name Lev Type
China Geological Hazard	ZSMA	Indicates rocky deserts and desertification.	Name Sarea Ftype Type
	ZNHP ZNHL ZNHA	Mudslides, landslides, avalanches, Collapse	Name Plac Lat Lon Dgrpop Dgrpro Stab Ftype Beleqk Eqklev

	ZDZP	The location where the earthquake occurred. Classification of earthquakes by magnitude	Name Eqktime Deep Eqklev Lat Lon
	ZDZA	Areas where earthquakes are likely to occur	Name Code
China National Nature Reserve	ZZMP ZZMA	Dotted with China's national nature reserves, China's national forest parks, China's national scenic spots, China's national geological parks, etc.	Name Dcode LEV Ftype Farea Proobj Belper Resfea Pubdat Admi
	ZZML	Regional boundaries of natural places of interest.	Name Dcode Lev Ftype
China Cultural Heritage	ZWYP ZWYA	Punctuated World Cultural and Natural Heritage, Chinese Tangible Cultural Heritage, Chinese Key Cultural Heritage Units, Chinese Intangible Cultural Heritage, etc.	Name Dcode LEV Ftype Farea Proobj Belper Resfea Pubdat Admi
	ZWYL	Regional boundaries of cultural heritage.	Name Dcode Lev Ftype
China Cultural Sites	ZWMP ZWMA	Dotted Chinese National Historical and Cultural Towns, Chinese National Historical and Cultural Villages, Chinese National Historical and Cultural Towns, Chinese Revolutionary Monuments, etc.	Name Dcode LEV Ftype Farea Proobj Belper Resfea Pubdat Admi
	ZWML	Regional boundaries of cultural attractions.	Name Dcode Lev Ftype

Table 1. Thematic Data Layer and Attribute Table Structure.

3.2 Thematic Map Database Building

According to the thematic map database design plan, relevant thematic information is collected and collated, and data analysis and processing are carried out to extract usable thematic information data through data normalisation and spatial processing. After strict quality checks and controls, a thematic data set that meets the requirements is formed and then spatially matched with the thematic geographic base map to form the thematic map database.

3.2.1 Thematic Map Database Building Process: In accordance with the database design methods and technical specifications, the thematic information and geographic base map information are first processed to generate thematic data sets and geographic base map data of qualified quality, and finally the thematic map database is established. The specific database building process is shown in Figure 1.

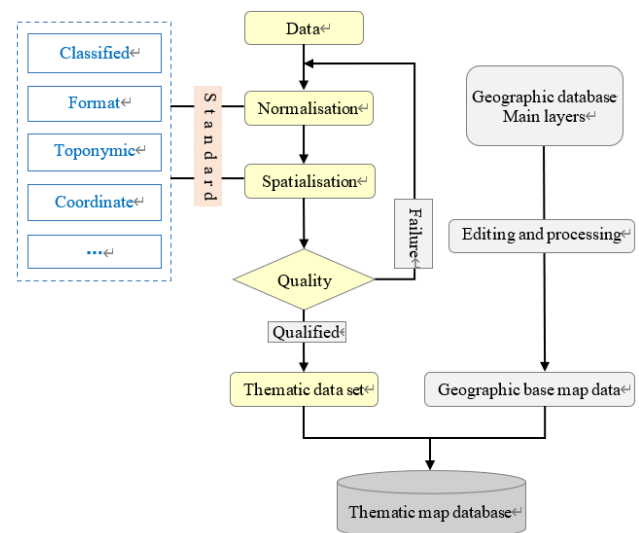


Figure 1. Thematic map database building process.

1) Production of thematic geographic base map data: Based on the main element layers in the latest general geographic map database, national basic geographic map data with 1:9 million map accuracy and comprehensive scale are produced, including basic elements such as water systems, inhabited areas, traffic, boundaries and place names, etc. Water systems indicate rivers of grade 4 or above and very large lakes, with appropriate rounding in dense areas; inhabited areas and their place names are indicated up to prefecture-level cities, with appropriate rounding; boundaries are indicated up to provincial boundaries. The boundary is indicated to the provincial boundary; transportation is indicated selectively according to the distribution characteristics of the whole country, such as high speed railways, double line railways, highways and national highways.

2) Thematic data collection and analysis: collect data for the selected topic of the thematic map, indicate the source of the data, and require authoritative and accurate data sources. The data collected in different formats and types will be classified according to the selected topic, screened, eliminated the worthless data, and organized and filed.

3) Normalisation of thematic data.

- Data format standardization processing: according to the data format, the data is divided into map data and information attribute data, of which map data is uniformly

processed into ArcGIS shapefile format, and information attribute data is uniformly transcribed into excel files, forming excel format files for storage.

- Place name normalisation: fields containing place names in the spatial map data and the excel data information attributes after collation are normalised according to the corresponding standards in order to be correctly associated with the standard place names in the basic geographical data. This includes adding common names to those without common names (adding common names such as "city, county, town, village, street"); changing simplified place names, especially those in ethnic minority areas, to full names and correcting incorrect place names; and correctly classifying place names according to those in the basic geographic data to improve the correctness and accuracy of later correlation operations, thereby enhancing the The work efficiency of data spatialisation.
- Coordinate normalisation: The coordinate system of spatial data is normalised in accordance with the standard "Standard representation of geographic point positions based on coordinates" (GB/T 16831-2013), with the following specific requirements: a sixty-increment system should be used for latitude and longitude; when the value of minutes or seconds is less than 10, 0 is added at the corresponding position; the units of degrees, minutes and seconds should be marked by symbols The recommended symbols are °, ' and " (see ISO/IEC 8859-1); the symbols should be located after the values; no space should be left between the values of degrees, minutes and seconds.

4) Spatialisation of thematic data: In order to match thematic map data to thematic geographical base maps, the following spatialisation processes are required.

- If the information collected is spatial map data, the information data is used directly as spatial data for the topic.
- The information collected is not spatial vector map data and needs to be placed in a base geographic information base and vector spatially corrected by reference points to make it thematic spatial data.
- The information collected is spatial raster map data, which needs to be corrected for spatial alignment, with projection settings, and then vectorised by software to become thematic spatial data.
- The information collected is information form data, which contains the geographic location coordinates values of the topic, can be directly converted into spatial point data through the existing ArcGIS function, and according to the specific expression content of the topic, the points will be connected to form line and surface data to form the spatial data of the topic; for the data of the topic that does not provide coordinates, it is necessary to create or improve the attribute field of the geographical name in the information form, associate it to the basic geographic data in the For thematic data where no coordinates are provided, the name attribute fields need to be created or refined in the information form, associated to the relevant layer in the basic geographic data, and extracted as thematic spatial data.

3.2.2 Technical Approach to Thematic Map Database

Building: Thematic map data building is the process of transforming the collected and collated thematic information into thematic map elements. The thematic element layer is determined according to the selected topic of the thematic map and the analysis of the collected information. On the basis of the basic geographic map database, point, line and surface elements that meet the expression requirements are added and the elements are layered according to the thematic content, the corresponding attribute fields are established, the field names, field types, field lengths and other related contents are clarified, and the collected thematic information is transferred into the layer attributes through association, merging and entry. The specific methods are as follows.

1) Extraction or association: The elements in the basic geographic map database are highlighted as thematic elements of the map sheets, such as (political areas, water systems, roads, railways, etc.). The corresponding topics in the basic database are extracted and the data of the collected and analysed thematic information forms are loaded into the extracted map elements of the basic database through the method of association, making it a thematic database.

2) Coordinate generation: Those generated by collecting coordinates provided in the data, and those re-corrected by raster graphics for vectorisation, need to be entered in the GIS as correlation items in their corresponding thematic information form data, and then the whole thematic information is loaded into the map elements by correlation to form a thematic library.

4. QUALITY REQUIREMENT

The quality of the sequence map thematic data directly affects the quality of the sequence map results of the next link, so it is necessary to strictly control the quality, to ensure the quality of the stage results, in order to ensure the quality of the final results.

4.1 Quality Control Content

Thematic data production and library building is an important part of this study, and is the basis for the development of the thematic maps of the sequence map, which requires high data quality. The design and research of the technical scheme in the previous phase lay the necessary technical foundation and quality guarantee for the subsequent data production and database building. Before data production, relevant technical personnel should be organised to study relevant standards, technical regulations and professional technical designs, and organise technical training to enable operators to properly understand, unify their understanding and proficiently master the technology. In the process of data production and database building, the subject implements the quality control mechanism of two-level inspection and one-level acceptance in accordance with the production operation process, the first-level inspection is the process quality inspection of the production results carried out by the quality inspection personnel of the production operation unit, the second-level inspection is the process quality inspection of the production results that have passed the first-level inspection by the quality inspection department of the production competent unit, and the first-level acceptance is the quality inspection of the production results that have passed the second-level inspection by the subject team. Level 1 acceptance

is the final quality acceptance of the production results by the subject team before the data is entered into the database.

In order to improve the quality of the sequence map thematic map and ensure the integrity, correctness and presentability of the results, comprehensive quality control and management of the spatial reference system, results specification, location accuracy, attribute accuracy, completeness, logical consistency, time accuracy and metadata quality of the results are required. Specific elements include the following.

1) Inspection of base map data: the basic data and the cartographic data should be inspected, and quality inspection and control is generally carried out in five aspects: mathematical accuracy, data and structural correctness, geographical accuracy, cartographic finishing quality and attachment quality. Mathematical accuracy, data and structural correctness, and geographical accuracy are the checks and controls on the quality of the basic data.

2) Checking the content of thematic data: mainly including the reasonableness of the classification and grading of thematic elements, the correctness of the representation method, the correctness of the relationship with other elements, the accuracy of various graphical data, etc.

4.2 Quality Control Methods

The technical methods and means used for checking the quality of the sequence map thematic data are a combination of automatic program checking, human-computer interaction checking and sample map proof testing, etc. The different checking methods have different advantages and need to be used in combination for practical applications.

1) Automatic program checking is the main focus

Using existing quality checking software, errors in the data are checked and found using the fact that there are certain logical relationships and patterns between graphs and attributes, graphs and graphs, and attributes and attributes of spatial data.

The method is fast and efficient and is used as the main method for checking the quality of data results. The disadvantage is that the correctness of automatic computer identification is not high enough and other methods are needed to assist.

2) Human-machine interaction check

The program check enables elements and locations that are in doubt to be searched out, narrowed down or pinpointed, and then the human-machine check method is used to determine the correctness of the data by hand.

This method has the best ratio of speed to correctness and is the main auxiliary method for quality checks. The method can be used for checking basic base map data results, thematic data results, etc.

3) Output sample check

Print out sample maps of the thematic data in the thematic map database and check the correctness of the features, data tables, or graphics by hand.

5. CONCLUSION

According to the design plan of the thematic map database, relevant thematic information is collected and collated, and data analysis and processing are carried out. The usable thematic information data is extracted through data normalisation and spatial processing, and after strict quality control, a thematic data set meeting the requirements is formed, and then spatially matched with the thematic geographic base map to form a thematic map database. In this paper, we have studied the processing of thematic data and the construction of a database for sequential maps, and have built a thematic map database that meets the requirements, based on which a thematic map of sequential maps is generated. This research result can provide reference and reference for other types of data processing and database construction.

ACKNOWLEDGEMENTS

This work is partly supported by National Key R&D Program of China (No. 2018YFC0807005).

REFERENCES

- CHEN, Y., 2005. A Research into Contemporary Atlas Design. Science Press, China.
- DU, X.R., TANG, J.J., and LIAO, J.J., 2011. China Atlas (2nd Edition). China Map Publishing House, China.
- LU, Y.S., 1997. National General Atlas Editing Design Features. *Cartography*, (04), pp. 22-24.
- National Geomatics Center of China, the editorial department of National Geomatics Atlas, 2015. The overall design scheme of new century version of the National Geomatics Atlas of the People's Republic of China.
- WANG, G.Z., WANG, D.H., ZHOU, W., CHEN, M., ZHAO, T.T., and FENG, Y., 2018a. Research And Compilation of City Maps in New Century Edition of "National Geomatics Atlas of the People's Republic of China". *Geomatics World*, 25(04), pp. 80-85.
- WANG, G.Z., WANG, D.H., ZHOU, W., CHEN, M., and ZHAO, T.T., 2018b. The Research and Compilation of City Maps in the National Geomatics Atlas of the People's Republic of China. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-3, 1703-1706. doi.org/10.5194/isprs-archives-XLII-3-1703-2018.
- ZHAO, T.T., and LIU, W.Z., 2018a. Research on the Refinement and Efficient Decision-making Map Compilation Based on MAPublisher. *Geomatics World*, 25(03), pp. 87-90.
- ZHAO, T.T., LIU, W.Z., and MA, W., 2018b. Research and Practice of the News Map Compilation Service. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-3, 2407-2410. doi.org/10.5194/isprs-archives-XLII-3-2407-2018.
- ZHAO, T.T., WANG, D.H., LIU, W.Z., ZHU, X.L., CHEN, M., WANG, G.Z., ZHOU, W., and MA, W., 2020. Research on

Thematic Maps in the New Century Edition of the National Geomatics Atlas of the People's Republic of China. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-3, 1077-1079. doi.org/10.5194/isprs-archives-XLII-3-W10-1077-2020.

ZHAO, T.T., LIU, W.Z., ZHU, X.L., WU, J., REN X.Y., PENG, Y.L., LI, R., and ZHAI, X., 2021. News Map Compilation and Application. *Geomatics World*, 2021. 28(1): p. 68-72.