Design and Engineering Practice of National Basic Geographic Entity Data Construction Technology

Yi ZHENG1
National Geomatics Center of China, Beijing 100830, China.

Abstract. The vector element data of traditional surveying and mapping production is a digital abstract expression of the real world in the form of 2 dimensions, it is difficult to fully meet the actual needs of analysis, calculation and use in the 3-dimensional digital space. In response to this issue, this paper takes China's 1:50000 fundamental geographic information feature data as the main data source to study the key technology and production process of national basic geographic entity data construction. Combined with engineering practice, the implementation process of key technologies such as transformation and reconstruction of geographic entities, semantic information supplement and improvement, entity coding, and entity relationship construction are described in detail. A practical and feasible process for constructing geographic entities has been formed, which effectively utilizes the advantages of geographic entities in information carrying, expression, sharing, and association. The first batch of national basic geographic entity data products show that the production process described in this article can quickly and effectively achieve the construction of geographic entities data, and can provide data and technical support for the integration of multi-scale geographic entities at the national, provincial and municipal levels.

Keywords. entity data, fundamental geographic information, entity coding, entity relationship construction

1. Introduction

Geographical entities are physical abstractions that humans use to describe and express geographical phenomena with specific spatial ranges, forms, processes, relationships, and related attributes in the geographical world. Their unique data form and the ability to effectively carry other professional information have the potential to integrate, share, and exchange information with various government departments[1].

Some developed countries in the international community have long carried out the physical integration and processing of basic geographic information data, forming a public geographic framework data that is easy to connect with thematic information, which can be regarded as the embryonic form of basic geographic entities[2]. In 1994, the U.S. government issued Executive Order 12906, requiring the establishment of the National Spatial Data Infrastructure (NSDI), and issued the "Geographic Information
Framework Data Content Series Standard”, which stipulated seven types of framework data, including cadastral, administrative management unit, water system, transportation, digital Orthophoto, elevation, and geodetic control, among which the first four types all use entity oriented data models[3]. In 2007, the European Parliament passed a resolution on building the European Union's spatial Information infrastructure (INSPIRE). Currently, INSPIRE has launched a series of data specifications. Similar to the framework data specification of FGDC, these data are also modeled as entity oriented as much as possible[4]. The Ordnance Survey of the UK began implementing the "Digital National Framework" program in 2000, building a new generation of data product MasterMap. Assign a unique and unchanging identification code TOID to all surface entities, and combine relevant entities through TOID to achieve the identification of geographical entities[5].

In recent years, in response to new development trends and strategic needs such as the construction of a digital China and the modernization of national spatial governance, many basic theoretical research and application explorations related to geographical entities have also been carried out in China. For example, Lu Guonian et al. described geographical entities from a geographical perspective and divided them into current geographical entities, planned geographical entities, constructed geographical entities, and managed geographical entities. They proposed granularity segmentation of geographical entities from the perspective of departmental applications[6]. Zhejiang Province has proposed the use of geographic entity encoding to achieve linkage update of multi-scale spatial data expression[7]. Jiangsu Province has conducted a unified classification and coding of all geographic entities and interest point information, exploring the use of geographic entity concepts to describe and express basic geographic information[8]. Wuhan, as an early pilot city, has designed geographical entity products and geographical scene products as new basic surveying and mapping products, and upgraded and transformed the urban basic geographic information database[9]. Shanghai is building a new type of basic geographic information spatiotemporal database based on geographical entities, taking into account the new demand for geographic information in the governance of super large cities and natural resource management[10].

The existing research and exploration of geographical entities at home and abroad are mainly limited to some types of land or local large-scale areas. However, the content, model structure, and application requirements at the macro level of national basic geographic entity data differ greatly from existing relevant research, and there is an urgent need to carry out technical design for constructing national basic geographic entity data[11]. This paper is mainly based on the existing 1:50000 basic geographical element data in China, combined with digital orthophoto map(DOM), digital elevation model(DEM) and relevant thematic data, relying on automatic transformation and reconstruction, semantic information supplement and improvement, entity semantic relationship construction and other technologies, to build national basic geographical entity data products, providing a basis for unified and standardized data linkage update and comprehensive information mining nationwide[12].

2. Overall technical approach

Following the overall approach of "utilizing inventory and introducing increment"[13], fully utilize existing basic surveying and mapping achievements, supplement and
improve relevant data content as needed, and carry out the construction of national level basic geographic entity data. Figure 1 shows the overall technical concept.

(1) Should be fully transferred. Elements in 1:50000 fundamental geographic information feature data with entity construction significance should be constructed into geographic entity, elements without entity construction significance are still saved as cartography features, ensuring the complete preservation of the content in 1:50000 fundamental geographic information feature data, and can be used to achieve cartographic representation based on basic geographic entity data.

(2) Seamless connection. Considering that 1:50000 fundamental geographic information feature data and basic geographic entity data will coexist for a period of time, and 1:50000 fundamental geographic information feature data is continuously updated, a one-to-one correspondence between 1:50000 fundamental geographic information feature data and basic geographic entity data through feature unique encoding (FEAID value) is established, ensuring that the updated 1:50000 fundamental geographic information feature data can be quickly linked to update basic geographic entity data in the future.

(3) Multi-granularity expression. The minimum granularity unit with entity significance generated by converting 1:50000 fundamental geographic information feature data is used as the basic granularity entity data at the national level, such as road sections, river sections, etc.

(4) Entity semantic relationships construction. Based on the Knowledge graph, the entity semantic relationship of basic geographical entity data is constructed to achieve the purpose of human-computer compatible understanding, analysis and application of
basic geographical entity data. For professional applications, the semantic relationship between basic geographic entities and professional entities can be expanded on this basis to meet the needs of industry sector applications[14].

3. Main technical processes

Using 1:50000 fundamental geographic information feature data, the entity transformation and reconstruction of elements with geographical entity significance are carried out, including data collection and analysis, data transformation, geographic entity coding, entity relationship construction, etc., ultimately forming the results of basic geographic entity data. Figure 2 shows the overall process.

Figure 2. Technical Flow Chart of National Basic Geographic Entity Data Construction.

(1) data collection and analysis. Collect and organize 1:50000 fundamental geographic information feature data, DOM for reference purposes, other natural
resource survey and monitoring data, and relevant thematic data, and analyze the
good, current situation, and availability of various types of data.

(2) data preprocessing. Perform format conversion, coordinate system conversion,
integrity check, etc. on the collected data and information.

(3) data conversion. It mainly includes three steps: establishing a basic geographic
tility data layer and attribute fields, creating a mapping table between features and
basic geographic entity data, and uniformly converting features into basic geographic
entity data according to the mapping relationship.

(4) data connection and fusion. Data connection and fusion should be carried out
between adjacent production units, including the data connection and fusion of basic
graphic entity data between adjacent production units within each production unit,
as well as between the data produced in the current year and the historical achievement
data.

(5) geographic entity code. Fill in the classification code, spatial identification
code and unique identification code of the basic geographic entity[15].

(6) entity relationship construction. Based on geographic entity data, combined
with images and other reference materials, construct entity semantic relationships,
mainly including combination relationship and association relationship between entities.

(7) metadata production. Create the required metadata.

(8) quality inspection. Perform quality inspection on basic data, geographic entity
data, metadata, etc.

4. Key technologies

4.1. Geographic Entity Data Transformation and Reconstruction Technology

It mainly includes the following parts:

(1) Construction of national basic geographic entity data model.

Create a national level basic geographic entity data template for storing
transformed entity data, combined entity data, and cartography features. The layer and
attribute information should meet the requirements of basic geographic entity data to
ensure the standardization of data storage after conversion. The attribute fields of the
entity dataset are composed of basic attribute fields and proprietary attribute fields. The
attribute fields of the cartography features dataset are consistent with the attribute
structure and spatial expression content of the 1:50000 fundamental geographic
information feature data fields.

(2) Mapping between 1:50000 fundamental geographic information feature data
and basic geographic entity.

Correspondingly convert the features with geographical entity significance in the
1:50000 fundamental geographic information feature data into natural geographic
elements, artificial geographic entities, and management geographic entities, while
retaining other features as cartography features without physical transformation.
According to the content of national basic geographic entity conversion, create a
mapping table between geographic information elements and basic geographic entities,
as well as a mapping table between geographic information elements and cartography
features, to standardize the correspondence between the graphics and attribute
information of geographic information elements and basic geographic entities during
the data conversion process.
(3) Automated transformation and reconstruction of entity data.

Carry out automated transformation and reconstruction work based on the basic geographic entity data layer and attribute requirements, as well as the mapping table between geographic information elements and basic geographic entities. Based on the created mapping table between geographic information elements and basic geographic entities, as well as the mapping table between geographic information elements and cartography features, the features that need to be materialized in the 1:50000 fundamental geographic information feature data are converted to the entity dataset, and the cartography features are also converted to the cartography features dataset.

4.2. Semantic Information Supplementation and Improvement Technology

It mainly includes the following parts:

(1) Obtaining semantic information based on identification information association.

By using unique identification information such as entity names and industry codes, geographic entity data is matched with thematic data to establish a mapping relationship between basic geographic entities and thematic data (such as road network, water network, etc.). Based on the mapping relationship between different data, the semantic information of geographic entities can be supplied and improved by incorporating the attributes contained in the thematic data.

(2) Obtaining semantic information based on spatial location association.

Unify the spatial coordinate system of geographic entity data and reference data. Establish a mapping relationship between geographic entities and reference data based on the matched spatial location information. Based on the mapping relationship between different data, the semantic information of geographic entities can be supplied and improved by incorporating the attributes contained in the thematic data[16].

(3) Obtaining semantic information based on semantic fusion.

Semantic fusion between geographic entity data and similar data in reference data. For example, the entity "river" in the water system and "river surface" in the reference material classification refer to the same type of geographical object. If a certain river object in the two types of data has the same spatial location information, then these two types of data represent the same geographical object, which can be semantically fused. The semantic information of geographic entities can be supplied and improved by incorporating the attributes contained in the semantic fused reference data.

4.3. Entity Semantic Relationship Construction Technology Based on Knowledge Graph

The types of entity semantic relationships mainly include combination relationship and association relationship between entities.

The entity combination relationship is the relationship between the combined entity and its constituent parts. Entity association relationships refer to various relationships between entities, including ownership, affiliation, adjacency, connection, passage, flow, and inflow. The entity relationships are sorted in order according to different categories (classification and coding order), and the same relationship is only recorded once between all two entities that have relationships. The entity relationship is recorded in the form of a relationship table and expressed in a "triple" format for standardization. National level basic geographic entities focus on constructing entity
semantic relationships for water systems, oceans, transportation, administrative regions, etc.

5. Engineering Implementation

In 2022, the Ministry of Natural Resources of the People’s Republic of China organized the National Geomatics Center of China, Shaanxi Bureau of Surveying, Mapping and Geoinformation, Heilongjiang Bureau of Surveying, Mapping and Geoinformation, and Sichuan Bureau of Surveying, Mapping and Geoinformation to carry out national basic geographic entity data trial production. The data achievements cover some areas of Qinghai and Gansu in the Yellow River basin, with a total area of about 202000km².

The national level basic geographic entity data mainly includes geographic entity types such as water systems, residential areas and facilities, transportation, pipelines, boundaries and administrative areas, natural and artificial landforms, and some vegetation and soil quality. Table 1 shows the specifications and indicators of its achievements.

<table>
<thead>
<tr>
<th>Data products</th>
<th>Data source</th>
<th>Mathematical foundations</th>
<th>Data format and specifications</th>
<th>Indicator requirements</th>
</tr>
</thead>
</table>
| National basic geographic entity data | Natural 1:50000 fundamental geographic information feature data, DOM, DEM, reference material | CGCS2000, 1985 National Elevation Datum | The entity data, relationship information and metadata are stored in vector files | Mean square error in plane: flat/hilly land 25meters
mountain/alpine 37.5meters
Mean square error of elevation: flat land 3meters
hilly land 5meters
mountain 8meters
alpine 14meters |

6. Conclusion

The construction of national basic geographic entity data is based on China's 1:50000 fundamental geographic information feature data as the main data source, utilizing technologies such as automated transformation and reconstruction, semantic information supplementation and improvement, and entity semantic relationship construction to produce basic geographic entity products with unified standards and specifications nationwide. It is an innovative utilization of existing national level basic geographic information data and an important practice to achieve the transformation and upgrading of surveying and mapping products from traditional element data to entity data. The construction of national level basic geographic entity data is an important task in the construction of a realistic 3D China, and also an important foundation for the future integration of multi-scale geographic entities in various provinces and cities.
References


