TECHNOLOGICAL FRAMEWORK OF INTELLIGENT MANAGEMENT AND APPLICATION ON GEOGRAPHIC INFORMATION

Ming Li [1,2], Shuliang Wang [1], Ju Bai [2] *, Xiaojia Liu [1], Chenzhe Wang [2]

[1] School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China - liming@ngcc.cn，(slwang2011, lxj) @bit.edu.cn

[2] National Geomatics Center of China, 28 Lianhuachi West Road, Beijing 100830, China - (baiju, wangchenzhe)@ngcc.cn

Commission IV, WG IV/7

KEY WORDS: Geographic Information, Intelligent Management, Quality Control, Life Cycle Intelligent Monitoring.

ABSTRACT:

The traditional geographic information management and application mode cannot meet the needs of intelligent management and on-demand application of massive multi-source heterogeneous data. Therefore, in this paper, we propose a technical framework of intelligent governance and application of geographic information. This technical framework is intended to realize the transformation of geographic information inspection from complexity to simplicity and stability, data management from isolated island to integration, and data application from single to on-demand. In the process of intelligent management of geographic information, this paper mainly carries out technical research from different links of collection, management, storage and apply, so as to optimize and improve the quality control, state detection, intelligent management and decision-making application of geographic information. The research contents include three parts, including quality intelligent control, multidimensional data management and life cycle intelligent monitoring.

1. INTRODUCTION

The geographic information is fundamental and strategic resource (Chen, 2010). The geographic information discussed in this paper mainly refers to the basic geographic data, including orthophoto data, topographic element data, digital elevation model data and topographic map mapping data (Tian et al., 2020). They are mainly composed of landform, water system, vegetation in natural geographic information and residential district, traffic, boundary, special features, place names and other elements in social geographic information (Jiang and Chen, 2000). The geographic information has the characteristics of multi carrier, multi space and multi time, which lead to the complex quality inspection steps, the difficulty of data sharing, and the simplification of data application (Chen et al., 2010). In order to improve the ability of production, processing and application, it is urgent to intelligently manage geographic information.

The quality of geographic data is the premise of data application guarantee, which directly affects the accuracy of decision-making based on geographic information (Zhang et al., 2015). Electronic document data management needs to meet the four characteristics, namely authenticity, integrity, availability and security (Wang, 2019). Therefore, the data archiving management department needs to control the quality of basic geographic data according to the four characteristics quality detection indicators and requirements of electronic documents. The quality evaluation of geographic information generally includes five links, including determining the applicable quality elements and sub elements, determining quality measurement indicators, designing reasonable and effective quality inspection methods, determining the quality evaluation rules and indicators, and evaluating geographic data (Wan, 2015). The research on the quality evaluation of basic geographic data began in the 1970s (Li et al., 2012). The research results mainly involve error source analysis, quality description and measurement, error propagation analysis, and quality evaluation and control. Data quality detection had been carried out by formulating standards (e.g. Specifications for quality inspection and acceptance of surveying and mapping products (GB/T 24356-2009) and Specifications for inspection and acceptance of quality of digital surveying and mapping achievements (GB/T 18316-2008)) (GB/T 24356-2009, 2009; GB/T 18316-2008, 2008) and designing quality detection models. However, most of quality detection models are only for a specific type or scale, and the inspection steps are complex, which directly affects the efficiency and increases the cost.

The basic geographic information usually takes the scale as the division standard to form data groups of different scales (Wang et al., 2018). The database of each scale includes multiple current versions, which can meet the multi-level needs of users' geographic information (Wang et al., 2013). To realize the intelligent management of multi-scale and multi-dimensional geographic data, it is necessary to master the status and attributes of data in each sub control link (data preparation, data receiving and inspection, data sorting and cataloguing, data management and database building, data query service and data application) (Peng et al., 2017). However, due to the problems of multi-source heterogeneity, cumbersome quality inspection and difficult traceability of geographic information, the existing spatio-temporal big data platform still has huge technical and management challenges in the management of geographic data (Devillers et al., 2010; Fan et al., 2010).

* Corresponding author

In summary, there are two difficult problems faced by geographic information management at this stage: one is that the quality inspection steps are cumbersome; the other one is that there are many data types, many intermediate achievement formats. The management of polymorphic storage media is missing. Therefore, in this paper, we propose a technical framework of intelligent governance and application to realize the following goals:
(1) Construct the four characteristics quantitative expression model to solve the problem of intelligent control of geographic data quality.
(2) Build a life cycle monitoring model to solve the problem of integrated intelligent management of data.
(3) Build a multidimensional geographic data management model to solve the problem of intelligent association management of geographic data.

The reminder of this paper is organized as follows: Section 2 presents the technical framework; Section 3 describes the implementation and evaluation; Conclusions and pointers to future work are given in Section 4.

2. TECHNICAL FRAMEWORK

Intelligent governance of geographic data refers to optimizing the quality control, state detection, intelligent management and decision-making application of basic geographic data. It mainly starts from the link of collection – management - storage - use, realizes the intelligent control of full quality characteristics, and intelligent monitoring of full life cycle.

In this paper, the geographic data has been classified, expanded and described, and then the quantitative expression model of geographic data is established to meet the requirements of descriptive information and quality control. The model will be expand into a multi-dimensional form + time + space data model to establish the intelligent association and management of basic geographic data with different storage carriers, different time series and different spatial scales. The data model is further expand as a data life cycle monitoring module to trace the source of polymorphic data. Figure 1 shows the overall process of establishing technical framework of intelligent governance and application.

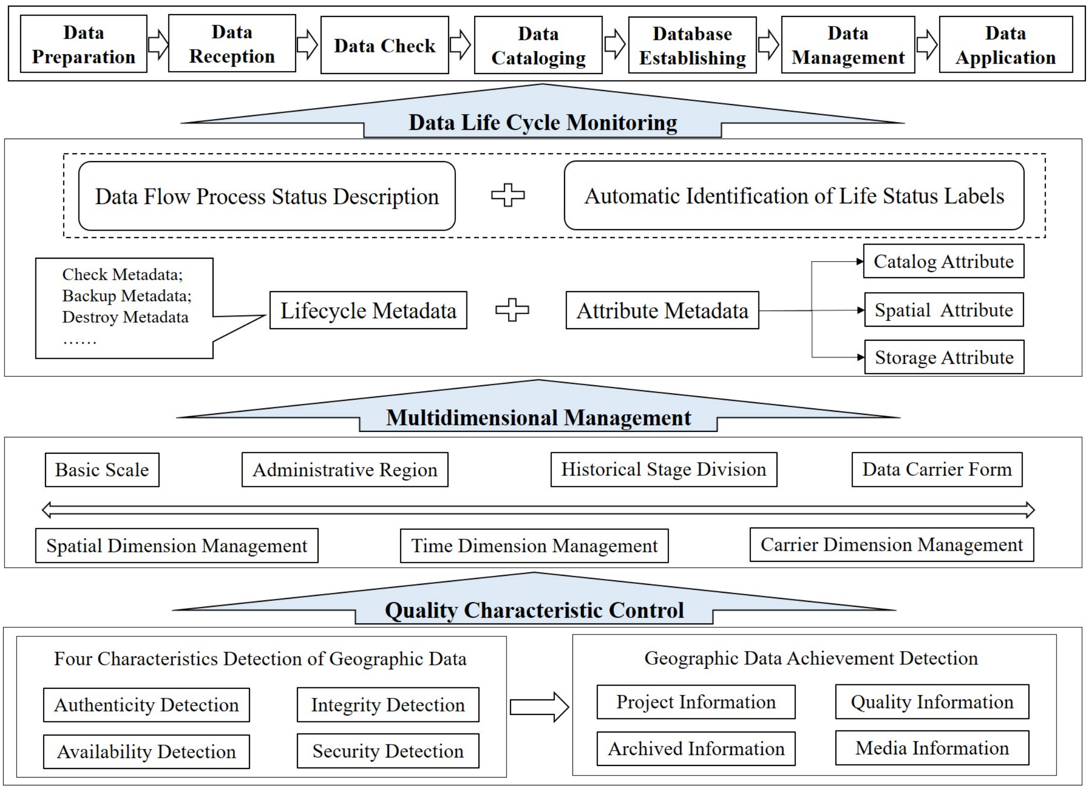


Figure 1. The technical framework of intelligent governance and application of geographic information.

2.1 Quality Characteristic Control

Before implementing geographic data into database, it is necessary to check the data quality to ensure the standardization and availability of data, to lay the foundation for subsequent data management and application. As described above, due to the variety and large amount of data, the traditional methods are often cumbersome and complex. In order to ensure the efficiency and quality of data quality inspection, an approach of classification, layering and grading by invoking different data inspection contents is adopted to complete the automatic detection and matching of multi-source basic geographic data.

As shown in figure 1, the quality characteristic control proposed in this paper includes two parts. One is four characteristics detection. The four characteristics of electronic documents are integrated into the quality control of geographic data, including authenticity detection, integrity detection, availability detection and security detection. It mainly detects the source of data, metadata, content, association between metadata and content, data volume, carrier, virus, etc. The other detection is to check up the data achievement, which needs to verify four aspects of information. Project information includes project metadata, total project amount, project identification code and project data amount. The archived information involves filing unit, filing metadata and filing identification code. Quality information refers to quality metadata, accuracy requirements and coverage integrity. Media information includes media metadata, backup unit and data confidentiality.

To realize intelligent data quality inspection, a four characteristics detection model of data is constructed. Firstly, the detection basis and method of each detection content is clarified. Then, each detection index is expressed quantitatively by using tamper proof technology. Finally, the model is optimized by autonomous learning algorithm, and then the intelligent semantic inspection and four characteristics quality control of basic geographic data are realized.

2.2 Multidimensional Geographic Data Management

In the stage of building basic geographic data database, we focus on the integrated management of multi-dimensional, multi period and multi state data. Firstly, a spatiotemporal integration reference datum characterized by uniform distribution, multi-resolution, multi-scale, seamless and open is established according to the national basic scale grid coding technology. It can be used to support the national multi-source heterogeneous data fusion across space-time, multi-scale and multi granularity. Secondly, through the combination of unstructured and structured basic geographic database technology, the storage space is unstructured expanded in the column dimension. This expansion should follow the scope of administrative divisions at national, provincial, municipal, county, and township. After extension, the user-defined description of the same source data is allowed. Thirdly, three technologies are proposed: automatic extraction of geospatial coordinates, coordinate association configuration of entity storage space matrix, and spatial matching Association of electronic data storage path. These technologies are used to solve the problem of multi spatial location matching of basic geographic data. Fourthly, based on spatial grid and administrative division, the spatial dimension, temporal dimension and morphological dimension of geographic data are further described and correlated. Finally, we build a multi-dimensional basic geographic data model of space dimension + time dimension + form dimension, and realize the multi-dimensional graphical expression and intelligent association management of basic geographic data.

2.3 Life Cycle Intelligent Monitoring

In the stage of basic geographic data management, the concept of life cycle is introduced into the management of basic geographic data. In the business process of geographic data, such as data preparation, reception and inspection, the description rule database and data format model database are established respectively. The two databases are built under the principle of Description rules for Surveying, Mapping and Geoinformation Records. The data types covered include surveying and mapping benchmark, aerial photography and satellite remote sensing, map mapping, remote sensing image data and thematic data. Through the automatic association and matching technology of data description information and format, the unified management of geographic data description attribute information, spatial attribute information and storage attribute information is realized. On this basis, an expression and trigger method of geographic data life state in each sub control link are studied. Through the automatic detection and identification technology of life state labels, to realize the intelligent monitoring of attribute information and life state attributes in the process of geographic data management.

3. IMPLEMENTATION AND EVALUATION

3.1 Prototype implementation

We have developed a prototype system of geographic information intelligent management to support Quality control, multi-dimensional management and life cycle monitoring of multi-source geographic data.

3.1.1 Quality control of archived data: The four characteristics checking includes the following steps.
Step1: check whether the content and quantity of archived data are correct and whether the data organization is standardized.
Step 2: check whether the archived data can be read normally.
Step 3: check whether the single image of raster data is correct and complete.
Step 4: check whether the layer content and structure of vector data unit are correct and complete.
Step 5: check whether the description content of the archived directory is correct and standardized.

This paper implements a set of methods for quality detection of raster data, vector data, GNSS (Global Navigation Satellite System) data and corresponding directory data. In order to realize intelligent detection, this paper uses item level, data type level and data format level to build index construction for subsequent automatic matching inspection. According to the established folder and file index information and the provided handover list keywords, automatically find the storage media information, stored folders and stored files with keywords. This paper puts forward the organization method of directory data file of surveying and mapping archives based on LINKID, which completes the association check of directory entities through digital matching and path matching. It is designed to quantitatively clarify the four characteristics detection content of various geographic data. Through the automatic comparison method of attribute identification code of digital archives, the quality inspection is completed.

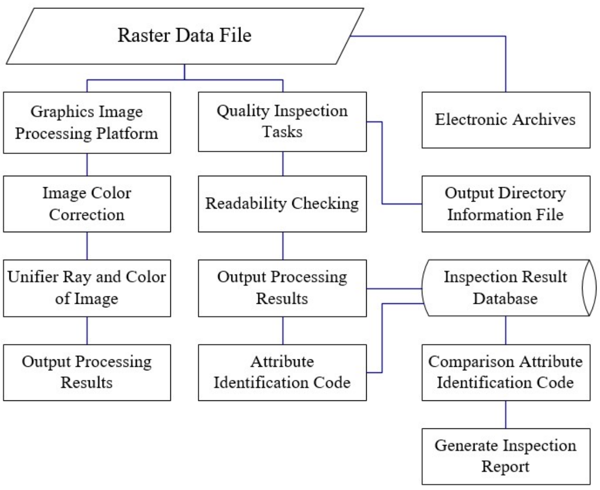
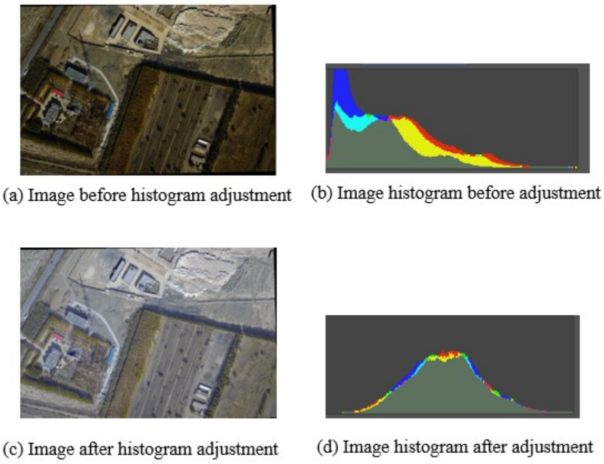


Figure 2. The quality inspection process of raster image.

Figure 2 shows the quality inspection process of raster image. There are two sources of remote sensing image: one is the digital original image or point cloud directly provided by aerial camera or satellite. The other is DOM (Digital orthophoto map) achievements formed from the production of Surveying and mapping projects. Usually, unifier ray and color of the original image plays a very important role for image quality. The detection of image color distribution features depends on calculating whether the image histogram conforms to the normal distribution. The histogram of standard image is calculated by gaussian formula, and it is used for histogram equalization.

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}(-\frac{(x-u)^2}{2\sigma^2}) \tag{1}$$

$f(x)$ represents the probability density function of the image at the x gray level, u is the mean value of all pixels of the image and σ is the standard deviation of all pixels of the image. Figure 3 shows the image effects before and after adjustment. After the above processing is completed, perform four characteristics checking and output the results. In order to realize the fast automatic quality inspection of raster image, this paper thins the image in proportion, and uses the image index map with small amount of data to quickly check the image situation and coverage.



(a) Image before histogram adjustment (b) Image histogram before adjustment

(c) Image after histogram adjustment (d) Image histogram after adjustment

Figure 3. Histogram adaptive adjustment.

3.1.2 Multidimensional Data Management:

(1) Extract the surveying and mapping archives directory data of all area graphics and carry out visual statistics and analysis directly in the form of map coverage. Therefore, users can quickly establish the overall view of data and comprehensively grasp the age coverage of data classification.

(2) Through decomposition and aggregation, data management from overall to detail is realized. It supports level by level decomposition and aggregation of classification years, so that business personnel can not only be sure of the whole, but also know a certain sub classification and a certain age range.

(3) Support fast data linkage by building graphic overlay cache. It can perform classification comparison, time comparison and administrative region decomposition, so that each data has actual graphic coverage support. Each coverage map can be compared with other coverage maps. Through mutual comparison and mutual verification, check and make up for omissions. Figure 4 shows a diagram of time-space dimension map coverage.

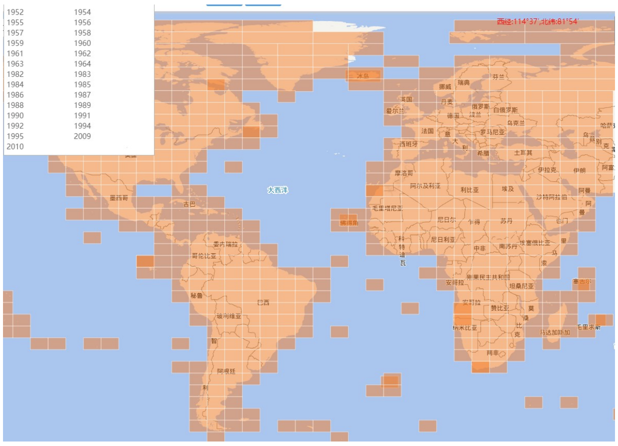


Figure 4. Time-space dimension map coverage diagram.

3.1.3 Data Flow Process Monitoring:

The seven links of geographic information management business include data preparation, data reception, data check, data cataloging, database establishing, data management and data application (figure 5), which can be summarized as collection, management, storage and use. They are linked together to form a closed loop of the whole life cycle of geographic data management. The core task of life cycle management is to realize the whole process information management from data warehousing to management and utilization. This requires the unification of various business models, attribute structures and association relationships involved in business processes. Transform the traditional management mode into dynamic management mode.

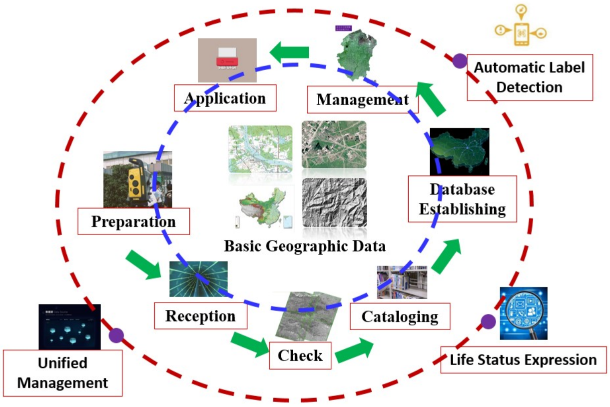


Figure 5. Geographic data lifecycle process.

After the data receiving being completed, life cycle management process is entered. The above seven links will be further refined into the following processes: data verification, data submission, distribution task, task receiving, simulation file composition, data grouping, description verification, primary quality inspection, integration submission, secondary quality inspection, integration review, warehousing handover and file warehousing. Each process will be registered and transferred to the next process after confirmation, ensuring the integrity of data life cycle management. The data entering the life cycle management process will automatically generate a unique tamper proof code, which will accompany the data for each process operation. Once the data changes, the code will change. By monitoring the tamper proof code, the status of the data in the whole life cycle management process can be monitored.

3.2 Evaluation

Generally, basic geographic information data include remote sensing image, point cloud data, grid data, vector data and database data. In terms of image files, it is necessary to judge whether the image file is valid by studying the reading methods of various types of image files, such as TIFF, JPG, IMG, ECW, raw, etc. Dilute the image in any proportion and save the thinned image to use the small image for quick view. In terms of point cloud data, it is necessary to judge data validity, including LAS, BIN and other data formats. The same is true for raster data, vector data and database data. They all need to judge the effectiveness of data according to their data format. Those qualified data will be archived. The validity check due to the diversity, complexity and mass of data commonly takes a long time by using traditional methods. The data quality control approach proposed in this paper can solve this problem.

To evaluate the performance of the method, several experiments are conducted with different processing threads. All experiments are carried out in a local area network (LAN) environment. We evaluate the quality inspection performance from time consumption. The experiment is to execute quality inspection using processing threads with an increased number (1 to 12). The experimental data include various data types. The data size is 80GB and contains 1703450 files. One personal computers (PC) is used, which is equipped with a 2.80 GHz Intel i7 processor and 32.0 GB of memory, running on Microsoft Windows 10.

The test is designed to be consistent with the scenario of daily data detection and warehousing. The test data are stored in 100 folders under the unified path. Test 1 uses the method proposed in this paper, and Test 2 simulates the traditional method. The environment and data source of the two experiments are completely consistent. It should be noted that traditional methods do not run in multiple threads. The time for the detection task execution is recorded each time and the average time is calculated from at least 10 executions. The evaluation results are shown in Table 1.

Test (Data Size:80 GB)	Test 1 time (unit: min)	Test 2 time (unit: min)
1 thread	40	65
2 threads	25	
3 threads	15	
4 threads	10	
5 threads	8	
6 threads	7	
7 threads	6	
8 threads	5	
9 threads	4.3	
10 threads	4	
11 threads	4	
12 threads	4	

Table 1. Performance test results.

Figure 6 shows the performance experimental results' tendency. The following observations can be noted:

(1) The time consumption of Test 1 is much lower than that of Test 2. The records in Table 1 also shows a similar conclusion. Within a certain range, the greater the number of threads deployed the better the detection approach performs. This shows that the method provided in this paper significantly improves the speed of geographic data quality detection.

(2) When the number of threads reaches ten, the time consumed is no longer reduced. The performance of thread pool depends on both the number of threads and the communication costs among them. When the number of thread reaches a certain number, the overhead of communications will have side effects on the performance. Thus, the growth will be slower than before.

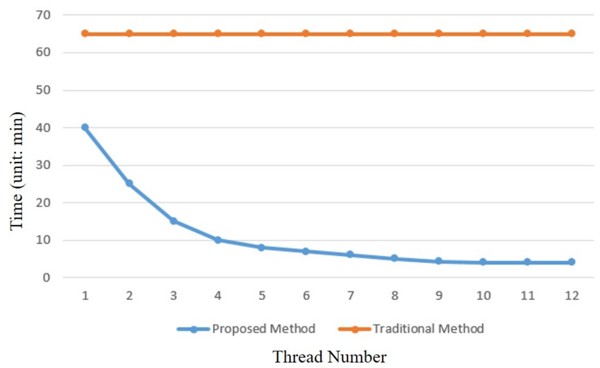


Figure 6. Performance evaluation results.

4. CONCLUSIONS AND FUTURE WORK

According to the different characteristics of multi carrier, multi space and multi time of basic geographic data, this paper constructs the four characteristics knowledge expression model of basic geographic data, and realizes the four characteristics intelligent semantic inspection of data. The data life cycle monitoring model is studied, and the whole life cycle intelligent governance system of basic geographic data is constructed. This paper studies the multi-dimensional organization method of basic geographic data, realizes the intelligent association of multi form, multi space and multi time data, solves the problem of cross departmental basic geographic data integration, and provides accurate, fine and high-quality basic geographic data guarantee for industrial applications.

ACKNOWLEDGEMENTS

The work is funded by National Natural Science Fund of China (62076027).

REFERENCES

Chen J. Construction of multidimensional dynamic geospatial framework data [J]. Journal of Geo-Information Science, 2012, 4(1):7-13.

Tian H.B, Chen L.J, Zhang H.W, Du X. Development of global 4D data products[J].Bulletin of Surveying and Mapping,2020(07):1-4+21.DOI:10.13474/j.cnki.11-2246.2020.0203.

Jiang J, Chen J. Thoughts on updating basic geographic information database [J]. Bulletin of Surveying and Mapping, 2000(5):3.

Chen J, Wang D.H, et al. Master Design and Technical Development for National 1：50000 Topographic Data-base Updating Engineering in China [J]. Acta Geodaetica et Cartographica Sinica, 2010(01):11-14.

Zhang L, Cheng P.F, Han W.L. Investigations of Quality Evaluation Standard for National Surveying and Mapping Key Project Products [J]. Bulletin of Surveying and Mapping, 2015(2).

Wang D. Interpretation of general requirements for inspection of document electronic archives [J]. China Archives, 2019, 546(04):29-30.

Wan Y.L. Research on the Theory for Spatial Data Quality Inspection and Assessment [D].Wuhan University,2015.

Li D , Zhang J , Wu H . Spatial data quality and beyond[J]. International Journal of Geographical Information Science, 2012, 26(12):2277-2290.

GB/T 24356-2009. Specifications for quality inspection and acceptance of surveying and mapping products [S]. 2009.

GB/T 18316-2008. Specifications for inspection and acceptance of quality of digital surveying and mapping achievements[S]. 2008.

Wang D.H, Liu J.J, Zhang Y,J, et al. Thoughts on Upgrading of National Fundamental Geographic Information Database [J]. Geomatics World, 2018, 025(002):1-6.

Wang S.L, Ding G.Y, Zhong M. On Spatial Data Mining under Big Data[J]. Journal of China Academy of Electronics and Information Technology, 2013(01):12-21.

Peng L, Jun Z, et al. The Research and Application of Basic Geographic Information Linkage Updating Technology[J]. geomatics & spatial information technology, 2017.

Devillers R , Stein A , Yvan Bédard, et al. Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities[J]. Transactions in GIS, 2010, 14(4):387-400.

Fan Y.T , Yang J.Y , Zhu D.H , et al. A time-based integration method of spatio-temporal data at spatial database level[J]. Mathematical and Computer Modelling, 2010, 51(11-12):1286-1292.