Database Construction and Application of Large-scale Geospatial Data

Zhaokun Zhai, Yin Gao, Jianjun Liu, Jianwei Liu, Jian Che

National Geomatics Center of China, Beijing, China - (zhaizhaokun, gaoyin, liujianjun, liujw, chejian)@ngcc.cn

Key Words: Geospatial Data, Database Construction, Data Management, Application System, Large Scale.

ABSTRACT:

Geospatial data refers to data that related to spatial location. It is the abstract representation of objects, events, or other characteristic information located on or near the surface of the Earth. It characterizes attribute features, spatial features, and associated information of geospatial features. Geospatial data commonly include various data types such as point, polyline, polygon, and so on. Geospatial database is the basic method to store and manage geospatial data. It is also the important component of geographic information systems (GIS). Geospatial database generally stores data in a specific structure in a medium, which can achieve efficient management of geospatial data. Users can add, modify, and delete data efficiently. Furthermore, query retrieval and statistical analysis can be quickly performed based on geospatial database. In this paper, distributed geospatial database is constructed by using database partition tables storage technology. Meanwhile, database construction software is developed to achieve the construction, management, and application of large-scale geospatial data. It can improve data import, query and retrieval efficiency significantly. The database application service system can directly connect to the database for displaying statistical analysis results. At the same time, geospatial data are processed to vector tiles, which stored in the tile database. Furthermore, standard online service is then published, which can be used for database management and application service system. It can be used for natural resource management work such as basic surveying, survey and monitoring, and so on.

1. Introduction

Geospatial data refers to data that related to spatial location. It is the abstract representation of objects, events, or other characteristic information located on or near the surface of the Earth (Gao, 2000). It characterizes attribute features, spatial features, and associated information of geospatial features. Geospatial data commonly include various data types such as point, polyline, polygon, and so on. Compared with non-spatial data, the most typical feature of geospatial data is the presence of spatial location. This is the main characteristic of geospatial data, including position, shape, and size of spatial features. In addition, geospatial data also has attribute features such as general and specialized attribute information that represent geospatial features (Sun et al., 2022) (Cui et al.,2007). At the same time, geospatial data also has topological features that describe the proximity and connectivity relationships between geospatial features. Geospatial data can be used in various fields, including government decision-making, navigation and surveying, environmental and resource management, urban planning and development, as well as emergency response and disaster management. The importance of geospatial data is its ability to provide visual representation of the world. It can help decision-makers make wise choices based on accurate and authentic information. In general, geospatial data has characteristics of large amount, complex structure and various relationships. There are certain difficulties in data storage and management (Zhao et al., 2020).

Geospatial database is the basic method to store and manage geospatial data. It is also the important component of geographic information systems (GIS) (Chen et al., 1999) (Wu et al., 2001). Geospatial database generally stores data in a specific structure in a medium, which can achieve efficient management of geospatial data (Liu et al., 2002). Users can add, modify, and delete data efficiently. Furthermore, query retrieval and statistical analysis can be quickly performed based on geospatial database. At the same time, geospatial databases have powerful geospatial analysis functions, which can perform spatial queries, spatial relationship analysis, spatial buffer analysis, and other operations. Geospatial analysis is one of the core functions of spatial databases, which can assist users in statistics, analysis, and modelling of geospatial data. Through geospatial analysis, users can reveal the inherent laws and correlations of geospatial data, providing scientific basis for decision-making and planning (Wang et al., 2022).

In this paper, distributed geospatial database is constructed by using database partition tables storage technology. Based on data content, data range, spatial reference, and data format of geospatial data, geospatial database design is done. Meanwhile, database construction software is developed to achieve the construction, management, and application of large-scale geospatial data (Zhu et al., 2004) (Chen et al., 2016). It mainly includes data inspection software and data import software, which can improve data quality check and import efficiency significantly (Liu et al., 2010) (Zeng et al., 2004). Then these geospatial data are checked by data quality software to ensure its quality (Duan al., 2022) (Zhang et al., 2020) (Zhai et al., 2021). These geospatial data are then imported into geospatial database by using data import software. The database management and application service system can directly connect to the database for displaying statistical analysis results. At the same time, geospatial data are processed to vector tiles, which stored in the tile database (Li et al., 2018) (Weng et al., 2018). Furthermore, standard online service is then published, which can be used for database management and application service system (Xia et al., 1997) (Chen et al., 2022). It can be used for natural resource management work such as basic surveying, survey and monitoring, and so on (Liu et al., 2021) (Liu et al., 2022).

2. General Design

Based on distributed database, a distributed geospatial database is constructed by database partition table technology to store and manage large-scale geospatial data. The efficiency of data import has significantly improved. Meanwhile, query

performance was also significantly enhanced, as well as the manageability of database tables, and the availability of data.

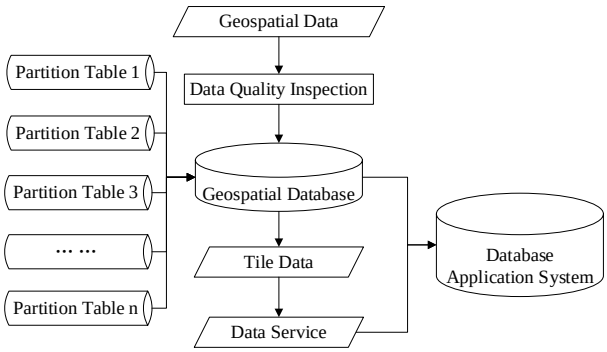

Figure 1. General Design of Database Construction and Application for Large-scale Geospatial Data.

Data analysis is finished to ensure spatial reference, data range, data content, data organization, data format, and technical requirements of geospatial data. Conceptual design, logical design, and physical design of geospatial databases are finished based on the content and technical requirements of geospatial data. Database software is developed to improve the work efficiency, including data inspection software and data import software. Data quality inspection is completed through automated data inspection software to ensure the quality of geospatial data.

Then geospatial database entities need to be built to distinguish different databases. Meanwhile, database table space is created to store geospatial data. Furthermore, in order to improve the efficiency of data import and management, database partition tables are used for storage geospatial data.

Geospatial data are then imported into its own database partition tables independently and parallelly, which can fully make full use of the performance of distributed databases. Consistency checks need to be done to ensure the completeness and correctness of data after the database construction is finished. Then database indexes are constructed to optimize database performance. Therefore, database application service system can directly connect to the database. At the same time, tile data are generated based on geospatial data, which can provide data support for database management and application systems to leverage the value of geospatial data.

3. Technical Method

3.1 Geospatial Data Analysis

To carry out database construction work, data analysis of geospatial data is necessary before database construction. It commonly includes spatial reference, data range, data content, data organization, data format, and technical requirements of geospatial data. Besides, the amount of data needs to be estimated in order to prepare data storage space reasonably. It should be emphasized that these data must meet the quality requirements of data production work.

The purpose of these work is to have a better understanding of the data, which is prepared for data database construction. Only with sufficient understanding of the data can we do a good job in the subsequent work.

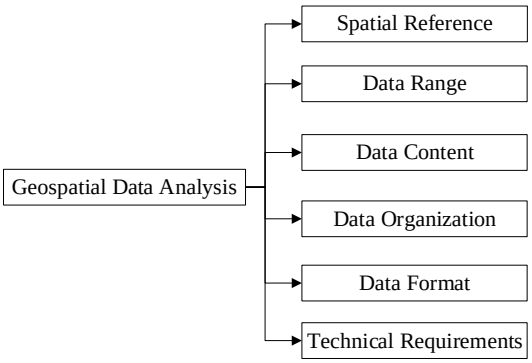

Figure 2. Geospatial Data Analysis.

3.2 General Structure of Geospatial Database System

Geospatial database system mainly consists of infrastructure, geospatial database, database management and application service system.

Infrastructure is the software, hardware, and network environment that supports the operation of the entire geospatial database, geospatial database management, and application service system. It mainly includes computer resources, storage resources, network resources, and security devices.

Geospatial database is the data resource of the entire geospatial database system, which can provide storage and management capabilities for geospatial data. Geospatial databases typically use relational databases such as Oracle Spatial, PostgreSQL etc.

Based on infrastructure and geospatial databases, a standard data access interface is adopted to publish these geospatial data as online data services. A database management and application service system are constructed to achieve functions such as browse, layer management, attribute query, and basic measurement. These services can be integrated into database management and application service system. Furthermore, the online data service can also be used in other relevant business systems.

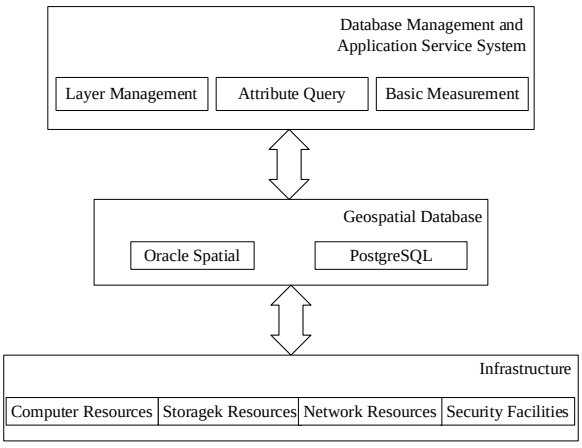

Figure 3. General Structure of Geospatial Database.

3.3 Geospatial Database Design

Adopting an overall approach of object-oriented approach, integrated management, and online services, geospatial database technology design should be finished, which can provide technical basis for geospatial database construction.

Based on contents and technical requirements of geospatial data, geospatial database design needs to be done, including concept design, logical design, physical design.

At this stage, it is important to clarify data that will be imported into geospatial database. At the same time, it is necessary to clarify the data content, definitions of data layer attribute fields, and data storage methods. The correlation between different layer needs to be analysed.

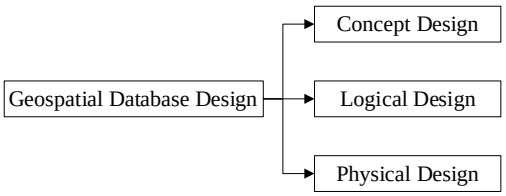

Figure 4. Geospatial Database Design.

3.4 Design and Development of Database Construction Software

To meet the needs of geospatial database construction and improve work efficiency, it is necessary to develop some related software, including data inspection software and data import software. Data inspection software can check the quality of geospatial data to ensure that these data can meet the quality requirements of database requirements. Data import software can import geospatial data into geospatial database efficiently.

In general, software has the characteristic of efficient processing, which can greatly improve database construction work efficiency. At the same time, it is also to reduce errors caused by manual processing. In order to achieve efficient work, multi-threading and multi-process technology is generally used during the work process.

The data inspection software can customize quality inspection plans based on data content and technical requirements. Meanwhile, the data import software can customize the import plan based on the data content. Both of these database construction software can achieve parallel processing of tasks.

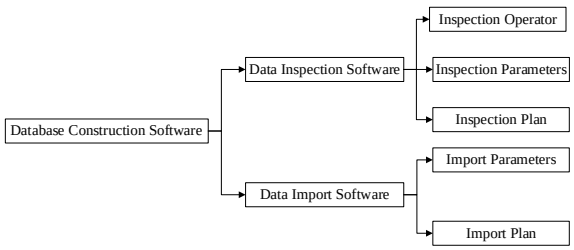

Figure 5. Design of Database Software.

3.5 Data Quality Inspection

The purpose of quality inspection is to ensure the quality of geospatial data. Meanwhile, this is also the foundation of the next work. All geospatial data must pass the quality inspection to ensure the smooth progress of subsequent database construction work.

After geospatial data summary is completed, quality inspection needs to be done to ensure data quality. This stage mainly focuses on the quality requirements for geospatial database. The inspection mainly includes file and structure consistency, spatial reference correctness, attribute consistency and topological consistency.

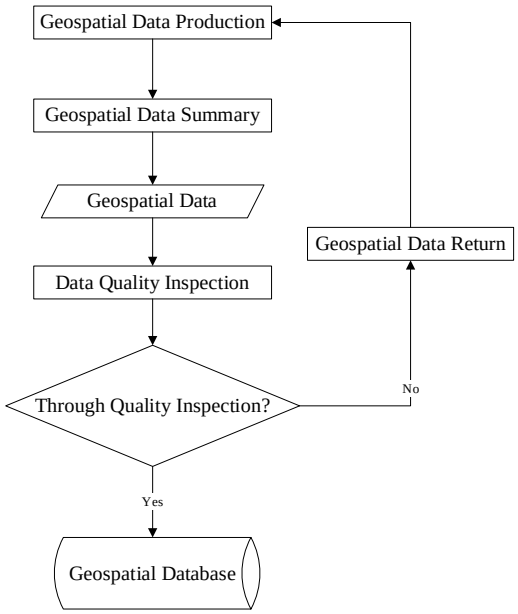

Figure 6. Data Quality Inspection.

Generally, the accuracy of data content that should be quality controlled during the production process of results is not checked.

If data quality issues are found during data quality inspection, these data will be returned to the data production unit. These data need to be reprocessed until these data meet quality requirements.

Finally, the data results that meet the quality requirements for storage will be summarized and organized to prepare for the next stage of storage work.

3.6 3.6 Geospatial Database Import

Geospatial database is the data resource of the entire database system, which can provide storage and management capabilities for geospatial data. Database import refers to importing geospatial data into a relational geospatial database. Based on data content and technology requirements, it is necessary to prepare for storage, such as creating corresponding tablespaces and database tables.

Partitioned tables can provide performance for data query and retrieval. Besides, partition tables are generally set based on the actual situation of geospatial data. If the amount of geospatial data is relatively large, partition tables need to be used to store geospatial data.

Considering the large amount of geospatial data, we generally use data import software to import data into geospatial database. By setting up the data entry plan and related parameters for the data import software, geospatial data can be automatically imported into a relational database.

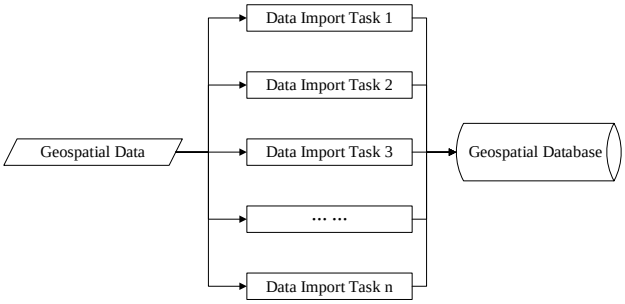

Figure 7. Geospatial Database Import.

After geospatial database construction work is completed, it is necessary to perform some checks based on the geospatial database to ensure the accuracy of data import work. The number of features between geospatial data and geospatial database should be consistent. If it is found that the number of features between geospatial data and geospatial database is inconsistent, the problem geospatial data needs to be reprocessed and imported.

It is necessary to check whether all types of data have been fully stored, including data range and data content. Meanwhile, it is also needed to check whether all types of data have been correctly stored, including names and attribute definitions of the dataset and data layers. Besides, the logical relationship of the database needed to be checked, including data storage logic, data partition logic, measurement area code logic, organizational structure logic. Meanwhile, attribute items of each data layer need to be checked to ensure the correctness of geospatial database.

To improve database quality and operational efficiency. After the completion of the achievement database construction work, further optimization work needs to be done on the geospatial database. Furthermore, indexes for geospatial data need to be done to achieve better performance, including attribute indexes and spatial indexes.

3.7 Online Data Service Publication

Nowadays, many businesses require real-time online data service. The performance requirements for online map services are also increasing. Fortunately, caching technology is a very effective way to improve the performance of data services, which can greatly reduce the pressure of server. Geospatial data are pre-processed into tile data Then geospatial data can be displayed and used by browser. Most importantly, there is no need to install professional software on client computer.

Map tile technology is commonly used in online map services. Tiles are the specific storage form of map data. In general, tile data are pre-processed in advance, which can greatly improve the access efficiency of online maps. For geospatial data, vector tiles are commonly used to display geospatial data, which has better performance.

After geospatial database construction work is completed, these data need to be processed into tile data. Then tile data are processed as a standard online data service, which can provide data services for database management systems.

If the geospatial database is updated, online data services should also be updated in time, which can improve the timeliness of online service data effectively.

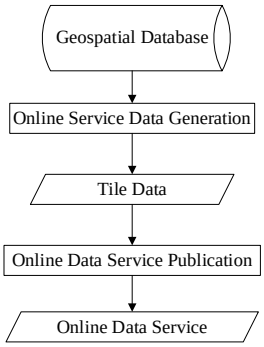

Figure 8. Online Data Service Publication.

3.8 Geospatial Database Management and Application System Integration

Geospatial database manage and application system is a system that can display and use geospatial data that managed by geospatial database. It commonly composed of hardware, software, database, and data administrators. It can manage and display the geospatial database uniformly to ensure its security and performance.

After the online data service is released, in order to better utilize geospatial data, online data services need to be integrated into the database management service system, which can realize unified management and display of geospatial data. Users can browse and use geospatial data in the geospatial database management system.

The geospatial database management system has basic functions, such as data browse and display, layer management, attribute query, coordinate positioning, basic measurement, and flight roaming. The database management system can integrate and display various geospatial data in the same scene. Meanwhile, corresponding application scene can also be developed for specific businesses.

Many daily works can be completed through geospatial database management system, which can enable collaboration among multiple people. Besides, it can greatly improve work efficiency.

4. Engineering Application

In this paper, taking the national geographic conditions monitoring data as an example, geospatial database construction work is completed. Furthermore, these geospatial data are processed to tile data. Online data service is then published, which can be integrated by geospatial database manage system. These data can be displayed and used in the same scene, which can greatly improve the efficiency. Users can also use these data via the browser conveniently, especially for uses without professional knowledge.

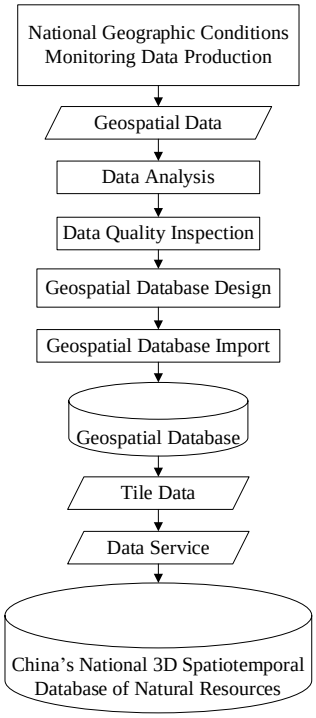

Figure 9. General Design of National Geographic Conditions Monitoring Database Construction and Application.

Firstly, national geographic conditions monitoring data are analysed. Spatial reference, data range, data content, data organization, data format, and technical requirements of national geographic conditions monitoring data are confirmed. The amount of data is also needed to estimate. This is the foundation of the next work.

Secondly, the design of national geographic conditions monitoring geospatial database is completed. It mainly includes conceptual design, logical design, physical design, and database software and hardware design.

The overall conceptual model of national geographic conditions monitoring geospatial database is shown in the following figure:

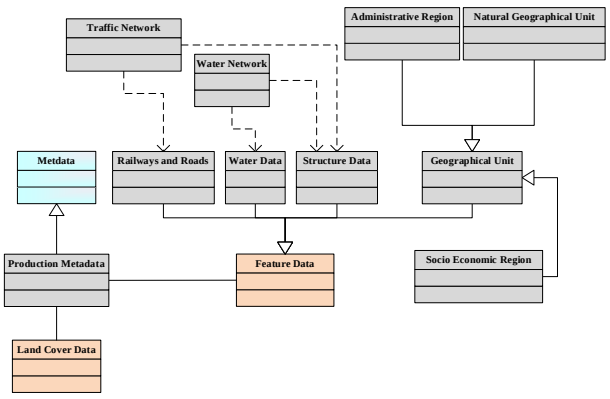

Figure 10. Overall Conceptual Model of National Geographic Conditions Monitoring Geospatial Database.

The overall logical structure of geographic national monitoring database is shown in the following figure:

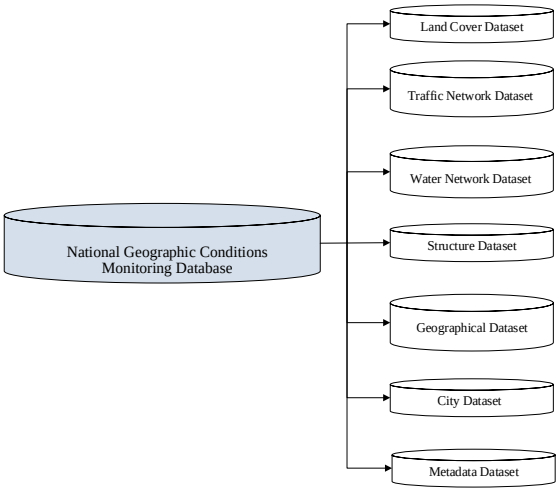

Figure 11. Overall Logical Structure of National Geographic National Monitoring Database.

Due to the large amount of national geographic conditions monitoring data, partition tables are used to store these data. The data that stored in each partition table is independent of each other. If a certain data needs to be replaced in the later stage, only this data needs to be replaced. The other data that stored in other partition table is not affected.

The partition table and its function of geographic national monitoring database is shown in the following table:

Table Space	Function	Remarks
GSurvey	Partition tablespace	Store national geographic conditions monitoring data without partition with provincial-level measurement areas as the storage granularity.
GSurvey_Partiton	Unpartitioned tablespace	Store national geographic conditions monitoring data that needs to be partitioned, with county-level measurement areas as the storage granularity.

Table 1. Functions of Tabel Space.

Thirdly, data quality inspection software and data import software are developed according to technical requirements of national geographic conditions monitoring geospatial data, which can greatly improve the efficiency of data quality inspection and data import.

Most importantly, data quality inspection software and data import software can also be used for other geospatial data by changing check scheme according to their check rules. It can meet the needs for data quality control of different projects. It reduces the cost and difficulty of later maintenance. It has high practicability and promotion value. The results show that the data quality inspection software based on check rules can improve data inspection efficiency and the quality of geospatial data.

Then national geographic conditions monitoring data can be checked efficiently by using automated quality inspection technology. The quality inspection of national geographic conditions monitoring data has been completed to ensure that these data can be imported into geospatial database.

After the quality inspection is completed, national geographic conditions monitoring data can be imported into the geospatial database. By using a multi terminal and multi-task parallel data import technology, national geographic conditions monitoring data are parallelly imported into relational databases, which can fully use the high concurrency performance of geospatial database and computers.

Based on high-performance vector tile technology, one click vector tile processing and publishing tool is developed, which can achieve rapid process of geospatial data. Therefore, national geographic conditions monitoring data is processed and published as an online data service. Then the data service is integrated into China’s National 3D Spatiotemporal Database of Natural Resources.

In this system, display, query, filter and render of geospatial can be done online. As the service is a standard data service, these online data services can also be integrated into relevant business systems. This may play a role in other business systems.

Figure 12. Integrated into China’s National 3D Spatiotemporal Database of Natural Resources.

5. Conclusions and Discussions

Currently, with the increasing accuracy of geospatial data collection and the great advancement of technological means, the amount of geospatial data is increasing. The difficulty of geospatial database construction and management is also increasing. At the same time, the application scenarios of geospatial data are also increasing, which traditional methods is no longer able to meet the need. Fortunately, distributed geospatial database can solve this problem.

In this paper, distributed geospatial database was constructed by using database partition tables. Therefore, management operations at the table level are dispersed to the partition level, and maintenance tasks are independently separated on each partition. Each partition table is independent of each other. When operating a partition, it does not affect the use of data from other partitions. Even if a partition is unavailable due to a failure, it will not affect application systems running on other partitions at all.

The overall data import efficiency is significantly improved. At the same time, the query performance of the database has been greatly improved. The manageability of the database tables and the availability of data have also been significantly improved.

Geospatial databases can be directly used by database management service systems. At the same time, these data can be processed to tile data. Then these data are published as standard data services. They can be used directly in application systems, which is efficient and flexible.

However, geospatial database management and application system running in the browser is unable to process large amounts of geospatial data directly. Fortunately, the geospatial database has the ability to process data, especially for large-scale geospatial data.

In the future, a new geospatial database management and application system is planned to develop. The database management and application system can publish data processing tasks in real-time. Then the geospatial database processes the corresponding data. Data processing results returns to the geospatial database management and application system. Therefore, users can use this system without professional knowledge.

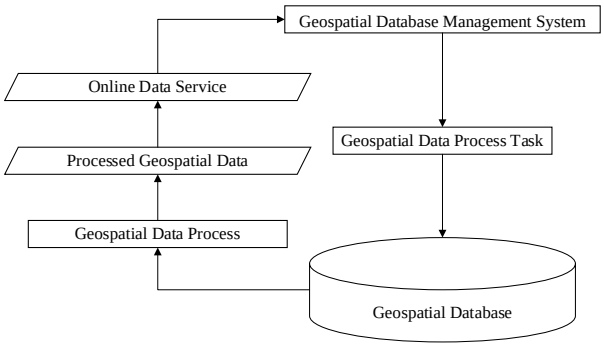

Figure 13. Flow Chart of New Geospatial Database Management and Application System

With the rapid development of the GIS, geographic information technology has changed from geographic information system to geographic information service (Gong et al., 2014). People hope to get knowledge from geospatial data directly. However, current geospatial data applications are mainly focused on providing data or data services. There is insufficient exploration of its knowledge stock. Very few geospatial knowledge services are provided (Chen et al., 2019).

Geospatial databases contain enormous data value. We should explore the value of geospatial data and provide geographic information services based on them, rather than just limited to providing geographic information data or data services.

Acknowledgements

I would like to thank DR Gao, and my colleagues for their previous efforts. The authors would also like to thank the National Geomatics Center of China for the support.

References

Chen F., Gong J. H., Chen L. Z., Meng Y., 2016. Design and Implementation of Quality Inspection System for Geographical Condition Census Based on Check Rules. *Bulletin of Surveying and Mapping*. 2016(3):122-125.

Chen J., Liu J. J., Tian H. B., 2022. Basic Directions and Technological Path for Building 3D Realistic Geospatial Scene in China. *Geomatics and Information Science of Wuhan University*, 47(10): 1568-1575.

Chen J., Liu W. Z., Wu H., Li Z. L., Zhao Y., Zhang L. 2019. Basic Issues and Research Agenda of Geospatial Knowledge Service. *Geomatics and Information Science of Wuhan University*, 44(1): 38-47.

Chen S. P., Lu X. J., 1999. *Introduction to Geographic Information Systems*. Science Press, Beijing, China.

Cui T. J., 2007. *Principles of Geospatial Database*. Science Press, Beijing, China.

Duan M. X., Li Y. D., Yan N., 2022. Design and application of geospatial data quality inspection platform. *Beijing Surveying and Mapping*. 36(11):1537-1541.

Gao J., 2000. Visualization in Geo-Spatial Data. *Engineering of Surveying and Mapping*, 9(3):1-7.

Gong J. Y., Geng J., Wu H. Y., 2014. Geospatial Knowledge Service:A Review. *Geomatics and Information Science of Wuhan University*, 39(8): 883-890.

Li H., Hu M. X., Gong Z. H., Fan B. J. 2018. 1Massive Spatial Data Parallel Processing Technology for Vector Tiles. *Computer and Modernization*. 2018(9):32-37.

Liu J. J., 2010. Study of Software Design for Quality Check of Fundamental Geographic Information Data. *Bulletin of Surveying and Mapping*., 2010(11):18-21.

Liu J. J., Liu J. W., Gao Y., Zhao W. H., Zhai Z. K., Che J., Zhao B. L., 2022. Technical Framework for Three-dimensional Spatio-temporal Main Database Construction of Natural Resources. *Geomatics World*, 29(3):37-4248.

Liu J. W., Gao Y., Zhao W. H.; Zhai Z. K., Wu C. C. 2021. Research on Construction Method of Natural Resources Spatial-temporal Data Model. *Geomatics World*, 28(5):42-46.

Liu N., Liu R. Y., 2002. *Geographic Information Systems*. Higher Education Press, Beijing, China.

Sun Q., Wen B. W., Chen X., 2022. Research on consistency processing of multi-source geospatial data. *Acta Geodaetica et Cartographica Sinica*, 51(7): 1561-1574.

Wang J. Y., 2022. Thoughts on the Future Development of Geographic Information System. *Geomatics and Information Science of Wuhan University*, 47(10): 1535-1545.

Weng S. J., Zhu X. J., Huang T., Ren F., 2018. Design and Implementation of Online Map Drawing Platform Based on Mapbox Vector Tile-sets. *Geomatics World*, 25(04):64-68.

Wu L., Liu Y., Zhang J., Ma X. C., Wei Z. Y., Tian Y., 2001. *Geographic Information Systems: Principles, Methods, and Applications*. Science Press, Beijing, China.

Xia X. K., Jia Q. R., Yang Q., Jin X., Li J., 2020. An Optimization Method of Spatial Data Loading for 3D WebGIS. *Geomatics and Information Science of Wuhan University*, 45(12): 1997-2004.

Zeng Y. W., Gong J. Y., 2004. Implementing Technique of Spatial Data Quality Control and Evaluation. *Geomatics and Information Science of Wuhan University*, 29(8): 686-690.

Zhai Z. K., Gao Y., Liu J. J., Zhang Y. J., Liu J. W., Zhang X. Q., Zhao W. H., Che J., 2021. Design and Development of Geospatial Data Quality Inspection Software Based on Check Rules. *Geomatics World*, 28(1):100-104.

Zhang B., Chen N., Gao Y., Pang X. F., 2020. The Design and Implementation of Vector Data Quality Inspection Tool. *Beijing Surveying and Mapping*. 34(10):1447-1450.

Zhao Y. P., Sun Q., Liu X. G., Cheng M. M., Yu T., Li Y. F., 2020. Geographical Entity-Oriented Semantic Similarity Measurement Method and Its Application in Road Matching. *Geomatics and Information Science of Wuhan University*, 45(5): 728-735.

Zhu Q., Chen S. L., Huang D., 2004. Key Issues on Quality Standardization of Geospatial Data. *Geomatics and Information Science of Wuhan University*, 29(10): 863-867.