Metadata Hierarchy in Integrated Geoscientific Database for Regional Mineral Prospecting

MA Xiaogang  WANG Xinqing  WU Chonglong  JU Feng

Abstract  One of the core developments in geomathematics in now days is the use of digital data processing in mineral prospecting and assessment. The information discovery is based on multidisciplinary geoscientific data and an integrated management approach is crucial. The lack of a standard description hinders interoperations in database search and discovery. Metadata hierarchy aims to provide a standard view of the geoscientific data, and facilitate data description and discovery. In the research of integrated geoscientific database, the metadata hierarchy used a standardized description for each collection in the content structure and realized in semantic structure. It recorded both dataset identification and inner structures and relationships of objects, thus differed from many other applications. There were four tiers in the content structure and three levels in the semantic structure. With its help, database users could determine how applicable a dataset is to a project, and improve their queries to the database. Effectiveness of data accessing is significantly enhanced through the rich, consistent metadata.

Keywords  metadata hierarchy; integrated geoscientific database; content structure; semantic structure

CLC number  P208

Introduction

Regional mineral assessment holds the knowledge-based data processing approach. Since the middle of 20th century, mineral prospecting has come into the so-called “scientific prospecting era”[1]. Computer-aided data processing and analyzing has been gradually taken into research and application[2]. It is a general way to carry out comprehensive analysis of geography, geology, mineral deposits, aeromagnetics, gravity, geochemistry and remote sensing datasets. Many scientists have developed specialized software, such as MORPAS[3], MRAS[4] and GeoDAS[5]. These applications require multidisciplinary datasets with consistent data structure and standard codes in the procedure of data processing.

Metadata acts as the analogue of library catalogs and thus aids the integration of datasets[6]. Some ongoing projects have carried out effective application of metadata in integrated geospatial databases. A typical example is the national geological map database[7] funded by USGS, in which the metadata conformed to the encoding guidelines established by FGDC. The structured metadata specified in formal terms both the syntactical structure and the semantic

Received on July 9, 2007.
Funded by the National 863 Program of China (No.2002AA130406) and the Key Project of China Geological Survey (No.200218310077).
MA Xiaogang, Institute of Land and Resources Information System, Faculty of Earth Resources, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China.
E-mail: maxiaog_cug@126.com
content, and properly described the digital geologic data\[8\]. Another example is the Geosciences Network (GEON)\[9\] funded by NSF of USA. The project team created a meaningful metadata model to incorporate with the job submission service and carried out a successful case study in synthetic seismogram generation\[10\]. Properly arranged metadata helps access through the integrated database and support the use of digital information.

As a representative metadata application for integrated geoscientific information management, the research aimed at setting up the metadata hierarchy and realizing integrated management of the multidisciplinary datasets from fourteen national mineral zones in China. Corresponding to the widely accepted three stages of resource discovery (overview, search and details), we identified the evolution of metadata’s function in data description and data discovery, and then configured the content structure and semantic structure of the metadata. The following sections present the background, method, process, and results of metadata hierarchy in the integrated geoscientific database.

1 Analysis and description of metadata hierarchy

1.1 Evolution of metadata’s functions

Metadata comes in a variety of schemes and varying levels of comprehension and complexity. One can view this variety in a continuum from the simple to complex. When configuring the metadata hierarchy, an evolitional process was related to the application of metadata for data description and data discovery.

The research utilized both means of metadata as mentioned above. Metadata was typically used for data description. It captured information about the salient features of the database content and then recorded them in standardized metadata elements. The resulting metadata represented the original database. The process of data discovery was optimized through the use of metadata records that provide structured representations of datasets\[11\]. Metadata schemes that intend to support the data discovery included more detailed elements which characterize the data resources. According to the information discovery approach in mineral assessment, as Fig.1 shows, meta-

data mainly supported raw data gathering and classification (Step 2) in conventional work. While the expanded metadata supported both Step 2 and Step 3, and in some respects, even the Step 4.

1.2 Content and semantic structure of metadata hierarchy

The concept model of the metadata hierarchy closely suited the inner relationships of the datasets. The digital datasets from fourteen mineral zones of China were gathered in fourteen self-contained aggregations. Each dataset contained seven subjects: geography, geology, mineral deposits, aeromagnetics, gravity, geochemistry and remote sensing. As spatial data, each subject contained several layers and each layer contained quite a few attribute fields. According to the content structure, we set up the metadata hierarchy for the integrated geoscientific database (Fig.2), which was suitable for each mineral zone’s dataset. The four content tiers respectively recorded the metadata of dataset, subject, layer and code. From the top to the bottom, the metadata was more and more
specific to the basic elements of the dataset. At the semantic structure, metadata was defined in three levels: sections, entities and elements. They defined the object classes in the integrated database and described their relationships. This architecture also accorded with the regulation of international standard ISO19115 and Chinese departmental standard TD/T1016-2003.

The functions of the metadata hierarchy expanded to data description, data discovery and retrieve, even data administration. Configuration of metadata depended on what function was wished to achieve. As Fig.3 shows, dataset and subject metadata were designed for data description. The former contained three entities: identification, contact information and key words. The latter contained four entities: coordinate system, content description, data quality and updating record. Layer and code metadata were designed for data discovery. The former contained one entity: layer catalog. The latter contained two entities: attribute codes list and records integrity. The main elements in each entity were presented in Fig.2. The metadata hierarchy referred to ISO19115 and TD/T1016-2003 in descriptive formulas and coding rules.

---

**Fig.2** Metadata hierarchy and its inner content structure and semantic structure in integrated geoscience database

**Fig.3** Implementation approach for metadata
1.3 Features and functionalities of metadata hierarchy

The metadata hierarchy acted an important role in the integrated management of multidisciplinary geospatial data. In the stage of database development and datasets uploading, the simultaneously inputted records of dataset and subject metadata were the rudder for dataset quality. In system application and maintenance, the records of layer and code metadata were the portal for data retrieval. In system updating, the whole metadata hierarchy was the basic evidence for referring. Concretely, the metadata hierarchy had such features as follows.

1) Supporting standard database design and data uploading. The structure and content of metadata suited the current standards and working manual issued by the government. The detailed definition of fields and codes in the metadata realized the standardization of data modes and codes inside the database.

2) Managing the database content, and helping users to understand the structure of the dataset. The records of the dataset and subject metadata were the detailed descriptions to the geospatial datasets. These records could be easily accessed through user interfaces and the user could get a general picture of the dataset rapidly.

3) Supporting the dynamic data model and increasing the flexibility of the system. For the complexity of geological data models, each department or person had his own particular aspect of research, and would distill different fields and codes to form special objects. The code metadata supplied flexible support to the dynamic data model.

4) Supporting the recovering of database framework. The metadata hierarchy recorded the database description and the objects, attributes, relationships of datasets on different level. Once any files of the database were destroyed, the metadata hierarchy could be used to recover the framework of the database.

2 Implementation and results

2.1 Implementation approach

We took above design into the development of integrated geoscientific database for fourteen national mineral zones in China. The implementation approach consisted of four stages in the main stream (Fig.3). In metadata hierarchy definition, we analyzed the application domain and objective user, then ascertained the data objects in the multidisciplinary datasets and clarified their inner relationships. The staged results were the content structure of the metadata hierarchy. In metadata arrangement and coding, we referred to the descriptive rules and coding language in current standards for geoinformation metadata, then set up the semantic structure and the catalog of units. After above work the contents and semantic structure of the metadata hierarchy were certain, and the detailed catalog for metadata elements was listed out. For the target differentiation between data description and data discovery, in the next stage the two parts were migrated into different developing environment to package the metadata units as different modules in the database. The database system was integrated with the data processing applications, such as MORPAS and MRAS. After accessing the metadata management module and finding the data elements concerned, user could distill part of the datasets directly from the database, and then transfer them to the data processing application for further work.

2.2 Results

The dataset and subject metadata were realized in digital sheets. They were released along with the integrated database. The metadata management module acted as the portal to the dataset objects and relationships in the integrated database, as well as the tool for database maintenance. User could browse the information of spatial datasets through the records of the metadata. As Fig.4 shows, the structured contents presented to the user included: the identification of the general information, contract information of data sources, key words sheet, coordinate system, dataset updating record, spatial data description and layer catalog.

The layer and code metadata were realized as administrative module in the integrated database. The terms and codes and their arrangement structures of the metadata module must be consistent with those of the datasets to be uploaded. Through the arrangement and coding of the seven subjects, layer catalogs, and
data items, the detailed description of the data structures and relationships was setup and listed as the sheets in the user interfaces. Fig.5 shows the portal for code metadata of geological map database. Similar interfaces were realized for the datasets of geography, mineral deposits, aeromagnetics, gravity, geochemistry and remote sensing.

3 Conclusions

Metadata plays a critical role in documenting and maintaining interrelationships of objects in a database, as well as in indicating the authenticity, structural and procedural integrity, and degree of completeness of information objects. The work in metadata hierarchy laid out a successful evidence for the functional evolution of metadata. The implementation in the integrated geoscientific database proved that it is a good way to utilize metadata to assist data management. Effectiveness of data accessing was significantly enhanced through the rich, consistent metadata. On a broader view, the natural progression of any organization is to move beyond the product and into value-add services.

References

[1] Zhao Pengda(2001) Theories and methods for mineral exploration[M]. Wuhan: China University of Geosciences Press (in Chinese)

[2] Zhao Pengda(2002) “Three component” quantitative resources prediction and assessments: theory and practice of digital mineral prospecting[J]. Earth Science-Journal of China University of Geosciences, 27(5):482-489 (in Chinese)

[3] Hu Guangdao, Chen Jianguo(1998) Design of the metallic mineral resources assessment and analysis system[J]. Geological Science and Technology Information, 17(1): 45-49 (in Chinese)

[4] Xiao Keyan, Zhang Xiaohua, Song Guoyao, et al.(1999) Development of GIS-based mineral resources assessment system[J]. Earth Science-Journal of China University of Geosciences, 24(5): 525-528 (in Chinese)

[5] Cheng Qiuming(2003) Fractal and multifractal modeling of hydrothermal mineral deposit spectrum: application to gold deposits in Abitibi Area, Ontario, Canada[J]. Journal of China University of Geosciences, 14(3): 199-206

[6] Loudon T V(2000) Geoscience after IT (Part L): adjusting the emerging information system to new technology[J]. Computers & Geosciences, 26 (special issue): 109-121

[7] Soller D R, Berg T M(1995) Developing the national geologic map database[J]. Geotimes, 40(6): 16-18

[8] Schweitzer P(1997) A strategy for creation of formal metadata[OL]. http://ngmdb.usgs.gov/Info/standards/metadata/meta_strat.html

[9] GEON Work Group(2006) GEON 2006 annual report: Cyberinfrastructure for the geosciences[OL]. http://www.geongrid.org/communications/annual_reports/GEONAnnual06.pdf

[10] Youn C, Kaiser T, Santini C, et al.(2006) Developing metadata services for grid enabling scientific applications[C]. The International Conference on Computational Science 2006 (ICCS 2006). Reading, UK

[11] Ma Xiaogang, Wu Chonglong, Wang Xinqing, et al.(2005) Centralized management approach and database development of multisource geoscientific information[C]. IAMG 2005 conference, Toronto, Canada