

# Geospatial data infrastructure: The development of metadata for geo-information in China

Baiquan Xu<sup>1</sup>, Shiqiang Yan<sup>1,2</sup>, Qianju Wang<sup>1</sup>, Jian Lian<sup>1</sup>, Xiaoping Wu<sup>1</sup>, and Keyong Ding<sup>1</sup>

<sup>1</sup>Development Research Centre of China Geological Survey, Beijing, China.

<sup>2</sup>School of Earth and Space Sciences, Peking University, Beijing, China.

E-mail: [xbaiquan@mail.cgs.gov.cn](mailto:xbaiquan@mail.cgs.gov.cn)

**Abstract:** Stores of geoscience records are in constant flux. These stores are continually added to by new information, ideas and data, which are frequently revised. The geoscience record is restrained by human thought and technology for handling information. Conventional methods strive, with limited success, to maintain geoscience records which are readily susceptible and renewable. The information system must adapt to the diversity of ideas and data in geoscience and their changes through time. In China, more than 400,000 types of important geological data are collected and produced in geological work during the last two decades, including oil, natural gas and marine data, mine exploration, geophysical, geochemical, remote sensing and important local geological survey and research reports. Numerous geospatial databases are formed and stored in National Geological Archives (NGA) with available formats of MapGIS, ArcGIS, ArcINFO, Metalfile, Raster, SQL Server, Access and JPEG. But there is no effective way to warrant that the quality of information is adequate in theory and practice for decision making. The need for fast, reliable, accurate and up-to-date information by providing the Geographic Information System (GIS) communities are becoming insistent for all geoinformation producers and users in China. Since 2010, a series of geoinformation projects have been carried out under the leadership of the Ministry of Land and Resources (MLR), including (1) Integration, update and maintenance of geoinformation databases; (2) Standards research on clusterization and industrialization of information services; (3) Platform construction of geological data sharing; (4) Construction of key borehole databases; (5) Product development of information services. “Nine-System” of the basic framework has been proposed for the development and improvement of the geospatial data infrastructure, which are focused on the construction of the cluster organization, cluster service, convergence, database, product, policy, technology, standard and infrastructure systems. The development of geoinformation stores and services put forward a need for Geospatial Data Infrastructure (GDI) in China. In this paper, some of the ideas envisaged into the development of metadata in China are discussed.

## 1. Introduction

In the last decade, many organizations have participated in a number of geoinformation initiatives in many developing countries. Geoinformation is vital for the sustainable development of our planet Earth, and is improving the national economic construction and natural resources exploitation. Besides, the application of geoinformation can benefit the strategic plans on environmental protection and social affairs. The Spatial Data Infrastructure (SDI) includes the networked geospatial databases, data



handling facilities, and the resources of technology, politics, human and economy that interactively connected between users and tools. The SDI promotes geospatial data sharing throughout all levels (local, regional or national level) of government, the private, non-profit sectors and academia<sup>[1, 2, 3]</sup>. It provides a structured system of practices and relationships between data producers and users, which underpins design, implementation and maintenance mechanisms for facilitating standardization, data sharing and accessing at affordable costs<sup>[2, 3]</sup>.

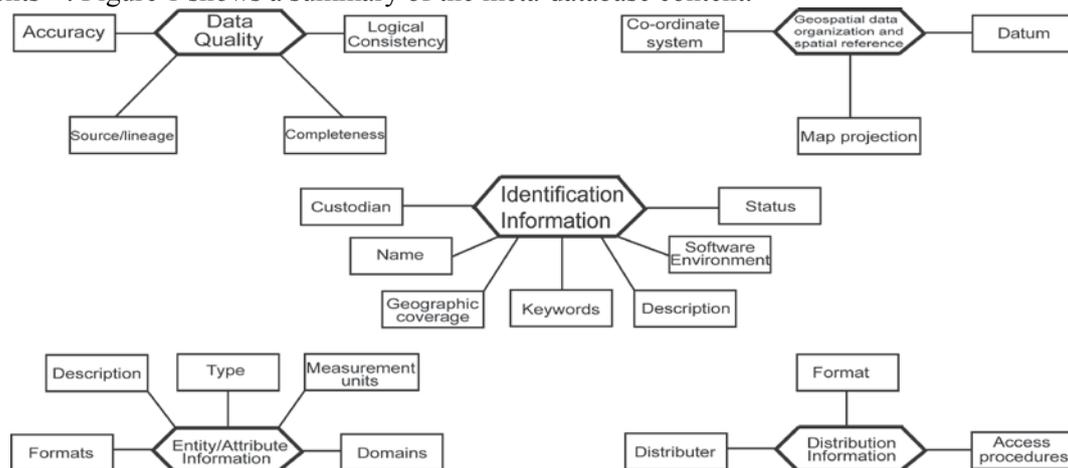
The Pre-1980 remote sensing technology in China was traditionally driven by downloading and storing data in computer. The fieldwork data was seldom delivered and transformed, which can hardly be traced by the data producer and end-users. The geosciences information had been collected with no geospatial reference and did not match with any standard data sets then. There was no connection between the geosciences data and end-users.

From 1980 to 2000, China began to carry out the digital transformation of the traditional geosciences data including the construction of the National Fundamental GIS (NFGIS) databases<sup>[4]</sup>, research on geo-information standards and digital survey of geo-mapping. These different geo-information databases were established on the basis of demand driven, application oriented and problem solving. However, the data was not well classified and the end-users were not clearly identified. There was a clear gap between the produced information and the absorbing of end-users.

From 2000 to present, the expansion of GIS has changed the traditional survey methods and has increased the storage capacity in the use of this kind of technology by various geo-survey institutes in China. Numerous geo-information and databases were available to the growing number of related public and national institutions for their research and decision making. The developing technology of Earth orbiting and observing provides a better resolution of the geospatial data to normal end-users who are interested in the geo-information for their own usage. The computer training helps both the researchers and citizens to gather local and international geo-information for their own needs.

## 2. Geospatial information management and data infrastructure in China

The proper documentations, including the content, quality, accessibility and other characteristics of the geospatial data, can enforce the need for good metadata and are meaningful to prospective end-users<sup>[3]</sup>. The International Organization for Standardization (ISO) developed a series of metadata standards and provided a schema for describing digital geo-data using a comprehensive set of conditional metadata elements<sup>[5]</sup>. Figure 1 shows a summary of the meta-database content.



**Figure 1.** Example of a meta-database content (modified from [3]).

The metadata elements are normally applied to the four major purposes as below:

**2.1 Discovery Identification** – The selected metadata elements enable the end-users to search the resources according to their needs. The metadata elements only cover simple information such as the author, title, subject, accuracy and the spatial reference system so as to limit the professional and specific searching enquires. Dublin Core is a typical representative.

2.2 *Cataloging* – Metadata elements, including the content, carrier, location, access, producing and using method of a dataset, are used for the detailed description on the comprehensive data units. Three of the typical representatives are MARC, GILS and FGDC .

2.3 *Resource Administration* – In addition to describing the comprehensive information, metadata elements support the information of data management by providing the privacy management, digital signature, seal of approval, access management, payment, accounting and more details of the information.

2.4 *Preservation and Archiving* – Metadata elements provide the information on the long-term preservation of the databases, including the data format, production information, protection condition, migration method and responsibility.

The core data of metadata is often produced, maintained, published and safeguarded by National Survey Agencies through consulting with the user communities. In general, governments of different levels (local, provincial and national) are in need of unlimited and efficient accesses to reliable, comprehensive and up-to-date geoinformation for their decision and strategy making. Stimulated by the progressive information technology, the development of geological exploration has entered a new period of digitalization and quantification. China started the digital transformation of traditional surveying and mapping technology system in the early 1980's, including the national fundamental geo-information system, application development and testing of basic geo-information standard system<sup>[6, 7]</sup>. The Ministry of Land and Resources (MLR) has made a plan for the development of geo-information system in China since 1986, including 12 sub-systems, 77 national databases and several model-and-method databases<sup>[8]</sup>. Over the past decades, plenty of fruits have been achieved in geological database construction, informatization of geological survey and exploration, domestic and international geological data collection, integration and public services.

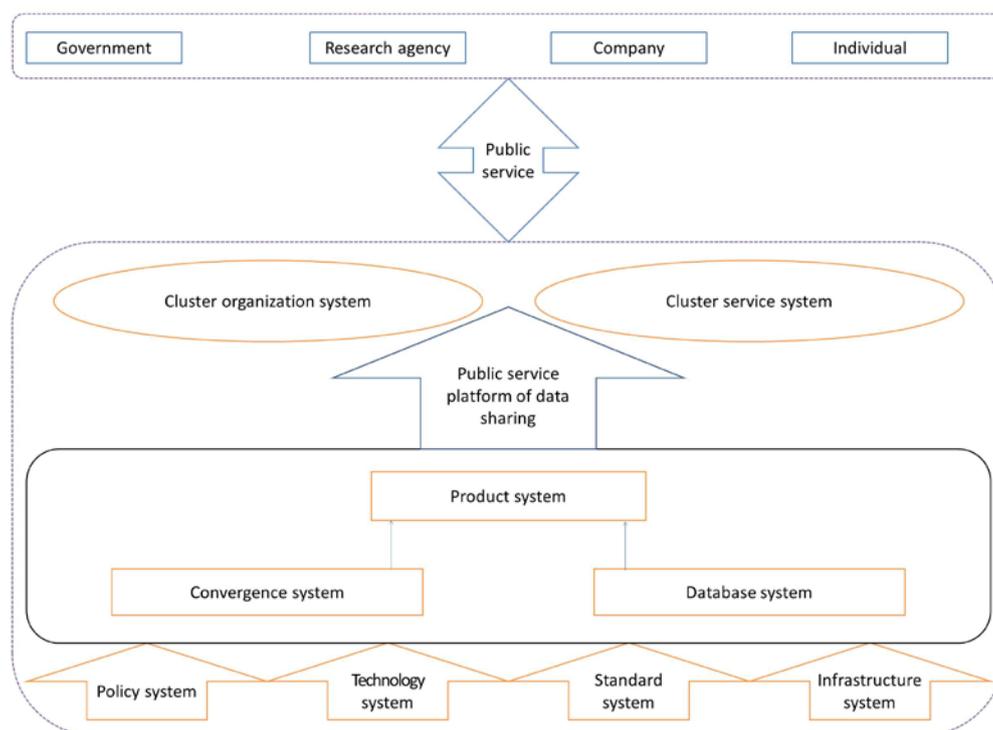
The National Geological Archives of China (NGAC) was officially founded in 2002, which is a government-sponsored organization under the auspices of the China Geological Survey (CGS) and MLR. The NGAC is in charge of the research on nationwide geo-data collection, management, maintenance and public services. It is also appointed by the CGS and MLR to provide technical support for national key projects. The government has been facilitating access and promoting the possible applications of fundamental geo-information in a well considered sharing and pricing policies, which include the obligation to produce a series of descriptions on the geo-data.

Since the 1980s', China has tried to maintain and manage the geo-information by dBASE, Basic/Fortran and AutoCAD for the mining exploration and geological mapping. Afterward, 3-D model became more popular in Chinese universities and institutes for their studies and researches. MapGIS platform is now widely used in geological survey and field mapping. Although these geo-information systems with independent copyright have the 3-D geological modeling and visualization functions, the development of geo-spatial information management in China is still in the beginning stage. The collected data is simple and not able to satisfy the multi-source geological data modeling requirements.

The Spatial Database Management System (SDMS) of CGS has been developed based on MAPGIS6.x SDK and Visual C++<sup>[9]</sup>, aiming at helping the government and the public to expediently use the spatial databases such as querying, indexing, mapping and product outputting. Now, the CGS has established 103 geological databases, and providing public services of 18 geological databases (Table 1)<sup>[10]</sup>. The spatial data in these databases were stored in multi-formats according to different managing software used, mainly in ArcInfo, ArcView (coverage or shape), MapGIS (a Chinese domestic GIS package, Wt, WP and WL), and MapInfo (Tab and Mif). Most of the database systems are developed on MS SQL server, some of MS Access and few of Oracle<sup>[11]</sup>. These databases will provide maps and final products of geological factors which are available to all the end-users according to the existing national laws and regulations in China. The users can submit their demands to the NGAC, and the NGAC will provide special hardware and software for the certain database processing.

**Table 1.** 18 geological databases provided for the public services in NGAC.

Name of the Database	Available Format
Geospatial Database of the 1:2 500 000 Geological map of China	MAPGIS, ARCINFO
Geospatial Database of the 1:5 000 000 Geological map of China	MAPGIS, ARCINFO
Aerp,agmetoc Database of the 1:5 000 000 Geological map of China	TM, ETM, MSS, CCD
Geospatial Database of the 1:6 000 000 Hydrological Geological map of China	MAPGIS, ARCINFO
China Geological Survey Extent Spatial Database	MAPGIS, ACCESS, SHAPE, JPEG
China Lithostratigraphic Database	DBF, SHAPE, ACCESS
China Geological Isotopic Dating Database	ACCESS
Thematic Map Geospatial Database of the China Geological Survey Deployment	MAPGIS, JPEG
Catalogue Database of the China Geological Data	ACCESS
China Regional Geochemistry Database	SQL SERVER, JPEG
Geospatial Database of the 1:500 000 Geological map of China	E00, WAT, WAL, WAP
Geospatial Database of the 1:200 000 Geological map of China	MAPGIS, ARCINFO, Metalfile, RASTER, JPEG
China Graph-Text Geological Data Database	JPEG, TIFF
China Mineral Product Place Database	ACCESS
Database of the 1:200 000 Native Heavy Concentration map of China	SQL SERVER, MAPGIS
Database of the 1:250 000 Aeromagnetic map of China	MAPGIS, JPEG
Geospatial Database of the 1:50 000 Geological map of China	MAPGIS, ARCGIS
Geospatial Database of the 1:200 000 Hydrological Geological map of China	MAPGIS, ARCINFO, Metalfile, RASTER, JPEG

**Figure 2.** The basic framework of clustering and industrialization of geo-data information services in China.

### 3. Geological data clusterization and industrialization in China

Metadata development is not only limited to the establishment of physical and organizational structures, but also the issues of data collection, delivery, exchange, share, publish and system maintenance. Each aspect of geological data management needs enormous financial and material supports.

A series of projects have been carried out since 2010, including the integration, regeneration and maintenance of the geological information databases; standard research on clusterization and industrialization of information services; platform construction of geological data sharing; construction of key borehole databases; product development of information services. “Nine-system” of the basic framework has been proposed for the development and improvement of the “clustering and industrialization of geo-data information services in China”, which is focused on the construction of the cluster organization, cluster service, convergence, database, product, policy, technology, standard and infrastructure systems (Figure 2).

*3.1 The cluster organization system* is mainly composed of governmental, industrial and business entities. The governmental organizations are responsible for the land and resources administrative management; the collection and public services of the geological data are managed by the national, provincial and commissioned organizations. The success of producing an efficient and reliable metadata depends on different entities, which are creating and using data and information. The Development Research Center (DRC) and six regional centers of the CGS are the main work forces for database development, regeneration and maintenance, computer network construction, and finally provide the professional services to the public. The data intermediary organizations and all levels of the geological prospecting organizations are in charge of commercial information services and constitute the organization systems of the business services. Now, there are 32 governmental administration organizations, 33 national and provincial data storage organizations, 35 data commissioned storage organizations and 6 professional servicing organizations in China. These organizations are responsible for collecting the geological data and providing public services. 35 commissioned storage organizations were commissioned by the MLR for storing oil and natural gas data in 2012.

*3.2 The cluster service system* is a 3-D network servicing pattern in conceptual design. Combining the conventional windowing with the modern networked servicing systems, the cluster service system is vertically consisted of integrated data services, regional data services, professional data services and primary data services, while horizontally of administrative public services, library public services, professional public services and commercial public services. Five geo-information servicing platforms are established in 2012 by the NGAC and provide more than 400 000 data catalogues for the public.

*3.3 The product system* involves a series of inter-connected servicing products including demand survey, data processing and system services. The product system research developed the basic theory of geological data production, the characteristics, factors and procedures of servicing products. The national and provincial geological data archives developed geological data searching, inquiring and browsing products. This helps the end-users find the products satisfying their needs. The NGAC compiled and provided the reliable, accurate and up-to-date geological survey maps for the governmental strategic planning and researches. The nationwide geological data integrated products supported the city construction and environmental protection, such as Shanghai 3-D city construction, Beijing ground subsidence monitoring, Anhui geological park planning and underground railway construction, 3-D digitalization of geological samples.

*3.4 The convergence system* is a standard mechanism for better submission and collection of the geological data. The MLR strengthened the geological data management through the departmental regulations. The primary data, final report and geological samples are entirely collected and centralized stored in national and provincial geological data archives. The MLR established a monitoring platform for the national geological data submission in 2012. This platform improved the quality of geological data submitting and accomplished the unified supervision of geological data.

*3.5 The database system* is a geological data integration and regeneration process. It includes the construction of regional geological survey, regional hydro-geological survey, marine geological survey, professional geological survey, exploration of mineral resources, water and environmental exploration, geophysical and geochemical survey, remote sensing imaging, thematic geological maps and other different types of databases. Now, the NGAC established 11 graphic and text databases, and set up a service-oriented database for promoting the integration of geological data.

3.6 *The policy system* is the roles for each participating organization in the establishing and subsequent maintenance of the metadata requires. To work in harmony and use effectively, users and coordinating entities in the governments and private individuals need a stable policy system. As the increasing needs for metadata, the MLR has established and released a series of policies for data collecting, sharing and publishing of geological data, right of the users, and data services.

3.7 *The technology system* refers to the facilities of data interpreting, processing, publishing and the methodologies of geological data surveying. The NGAC integrated and exploited the number of geological information and developed a standard technology system for data collecting, mining, exchanging, processing and sharing processes. The users can obtain their information from the NGAC by the traditional, online and newly developed mobile services.

3.8 *The standard system* is a framework of providing common sets of terminology and definition for documentation of spatial data. As the standards of different nations must be consistent with those in related fields, standardization of data is a very serious problem. There need to be a common working ground to stand on for different users to cooperate and achieve their objectives. The standards of China geological data informationalization include fundamental standards, database standards and specifications of software system development. To support the databases development and informationalization, a series of standards have been issued since 1999<sup>[11]</sup>. For spatial databases, CGS adopts uniform national geographic space reference coordinates system<sup>[11]</sup>. Nowadays, China has established a servicing standard for geological information system framework, especially for the servicing description and searching protocol in NGAC<sup>[10]</sup>.

3.9 *The infrastructure system* concerns the networked geospatial databases and data handling facilities. A proper metadata requires a solid infrastructure of policy, administrative arrangement, technical standards, and fundamental databases. As the clusterization and industrialization of information services processing, national and provincial geological data centers have been constructed since 2010. In the meanwhile, the infrastructure system construction was written in the national geosciences data servicing project.,

#### 4. Conclusion

The geo-spatial data infrastructure in China is in its infancy. Most of the data environment is still application-oriented. There are no consolidated and efficient systems to maintain, manage and utilize the geospatial databases. The available legacy databases suffer from inconsistency of the data storing, duplication technique and digital transformation. The construction of geological data system is built on the read-made commercial software, and is lack of secondary innovation. A national level of geo-information managing, accessing, and producing system should be composed of organizations and loosely connected users. The metadata and data exchange models are thus become crucial to Chinese database management. The implementation of a fixable metadata requires a solid infrastructure system, which is based on the governmental-level policy, guideline and administrative arrangement, international consistent technical standards, fundamental databases and accessible geo-spatial data communities.

#### References:

- [1] NSDI Future Directions Initiative: *Towards a National Geospatial Strategy and Implementation Plan*: <http://www.fgdc.gov/nsdi/nsdi.html/>.
- [2] Groot, R. and McLaughlin, J., (editors), 2000. *Geospatial Data Infrastructure: Concepts, cases and good practice*. Oxford University Press.
- [3] Woldai, T., 2002. *Geospatial data infrastructure : the problem of developing metadata for geoinformation in Africa*. In: Proceedings of the 4th International conference of the African Association of Remote Sensing of the Environment AARSE : Geoinformation for sustainable development in Africa : Abuja, Nigeria, 14-18 October 2002. 15 p.
- [4] Yongchang Yu, 1999. The buildup of China National spatial data infrastructure. *Research on Soft Science of Surveying and Mapping*, **2**, 1999, pp.9-14 (in Chinese).

- [5] International Organization of Standardization Homepage. <http://www.iso.org>.
- [6] Yangying Xu, 2000. Advances in China National spatial data infrastructure. *China Surveying and Mapping*, **4**, 2000, pp. 6-8 (in Chinese).
- [7] Deren Li, Jianya Gong, Jingwei Li and Mi Wang, 2002. Establishment of the spatial data infrastructure of China. *Bulletin of Surveying and Mapping*, **11**, 2002, pp. 4-7(in Chinese).
- [8] Chonglong, Wu, Computing technology in the field of geology and mineral resource. *Earth Science Frontiers*, **5**, 1998, pp.343-355.
- [9] Jianguo Chen, Zhijun Chen, Quanming Wang and Yiping Fang, 2003. Spatial database management system of China Geological Survey Extent. *Journal of China University of Geosciences*, **14 No.3**, 2003, pp.250-256.
- [10] National Geological Archives of China Homepage. <http://new.ngac.cn>.
- [11] Minghua Zhang and Qiherige, 2004. *Geo-information work at China Geological Survey*. Map Asia Conference 2004. Beijing, China.