

# Development on management and sharing of data collected in Chinese Antarctic and Arctic Research Expeditions

Ling Xiaoliang(凌晓良), Yang Bin(杨滨), Zhang Xia(张侠), Zhang Jie(张洁), Zhu Jiangang(朱建钢) and Wang Dali(汪大立)

*Polar Research Institute of China, Shanghai 200136, China*

Received May 25, 2004

**Abstract** Combined with the current status of Antarctic data management and the characteristics of polar science data resulted from Chinese Antarctic and Arctic Research Expeditions, the Chinese Polar Science Database System (CPSDS) has been designed and established in 2002. The infrastructure, technical standard, mechanism of sharing data of this system are reviewed in this article. Meanwhile, the development of Chinese polar data management is summarized. As the metadata is the powerful and useful tool for managing and disseminating scientific data, the metadata is also used as “search engine” of CPSDS. Besides, the trend of data management and sharing is also discussed.

**Key words** data management, metadata, database system.

## 1 Preface

With the rapid development of remote sensing technology, Internet technology and database technology, data acquisition, data processing, data storage and distribution in polar research or other research form a stable data-flow link. It is our common responsibility and objective to make the historical data from polar region easily available to all the scientists around the world, as Article III.1.c of the Antarctic Treaty states that Antarctic scientific data “should be fully and freely exchanged”. Obviously, it is an important and promising work to index, store, disseminate, link and mine the scientific data from Chinese Arctic and Antarctic research activities more effectively.

From 1999 through 2004, Chinese National Antarctic Data Center (CN-NADC), as a member of SCAR-COMNAP-JCADM, has established a web-based comprehensive database system, consisting of 8 different disciplinary sub-database, with the aim of integrating all raw data from Chinese Arctic and Antarctic research. During this process, the first step is

to retrieve, sort out and backup all data, most of which is held by scientists, data collector, universities and research institutions respectively, the second step is to index, catalog and publish data by linking the metadata and its dataset.

From now on, some of polar scientific data can be accessed from the Chinese Polar Scientific Database System (CN-PolarDB) via Internet, into which more than 1900 Antarctic datasets with the size of 80 GB, mostly of polar oceanography and solar-terrestrial physics, is now input. Besides, 54 datasets collected in the 2nd CHINARE-Arctic stored at CN-PolarDB is also available for registered users. As the national data strategic policy has been defined and its regulation came in force in 2002, the scientists anticipating the CHINARE and CHINARE-Arctic are required to submit or update data (including metadata) in an acceptable form to the CN-NADC within a period of time. This year, the data concerning the polar glaciology and meteorology is scheduled to integrate into the CN-PolarDB.

## 2 The Chinese Polar Scientific Database System

CN-PolarDB is a web-based comprehensive database of scientific data, in which the data from survey and sample analysis, linked with their respective metadata, can be stored and accessed in a general format, searched and downloaded freely.

### 2.1 *The introduction on the CHINARE and CHINARE-Arctic and polar data resources*

CN-PolarDB is composed of a broad range of scientific data acquired in CHINARE and CHINARE-Arctic, either raw or processed in any format, including the long-term monitoring data collected on Chinese Antarctic and Arctic bases.

From 1984 to 2004, China has organized formally 20 expeditions to Antarctica (including 4 traverse Expeditions from Zhongshan station to Dome A), 2 expeditions to Arctic ocean, and 1 expedition to Arctic. Up to now, three bases (Chinese Antarctic Great Wall station, Chinese Antarctic Zhongshan Station, Chinese Arctic Huanghe Station) were established respectively in 1985, 1989, 2004. Up to now, a series of science projects was implemented, lots of data involving over ten disciplines—glaciology, geography, geology, geophysics, geodesy, oceanography, atmospheric science, meteorology, biology, astronomy and human medicine was acquired. The Regional Earth Observation System (REOS) located at Arctic and Antarctic regions has been so far established after the efforts of many years. It consists of some core observation sites: 2 Antarctic All-year stations, 1 Arctic All-year station, Xuelong Scientific Expedition Vessel, Traverse-Antarctica route.

To retrieve, backup and integrate and publish all the historical data, most of them still held respectively by polar scientists, data collectors, related universities and research insti-

tutions is an important urgent and hard work. Now we completed or are doing these tasks.

1): 1900 datasets, acquired from 1984 to 1999, concerning oceanography, solar-terrestrial physics and glaciology, has been retrieved and backuped at CN-NADC, with the data of 80 GB available via Internet.

2): Retrieving and back-up'ing all datasets acquired after 2000, namely 18th-21st CH NARE and 2nd CH NARE-Arctic

3): By 2008, to Retrieve and back-up 1500 datasets, acquired from 1985 to 1999, concerning geology, geo-physics, atmospheric science, geodesy, geography, biology, geog-raphy.

## 2.2 Infrastructure of Database System

The hardware platform of CN-PolarDB is based on general two-layer structure, consisting of web server and database server. Web server is responsible for processing the request from WWW users, and transforming it into database server. Application server runs only with local network, based on the Client/Server model, for direct data access such as adding, updating.

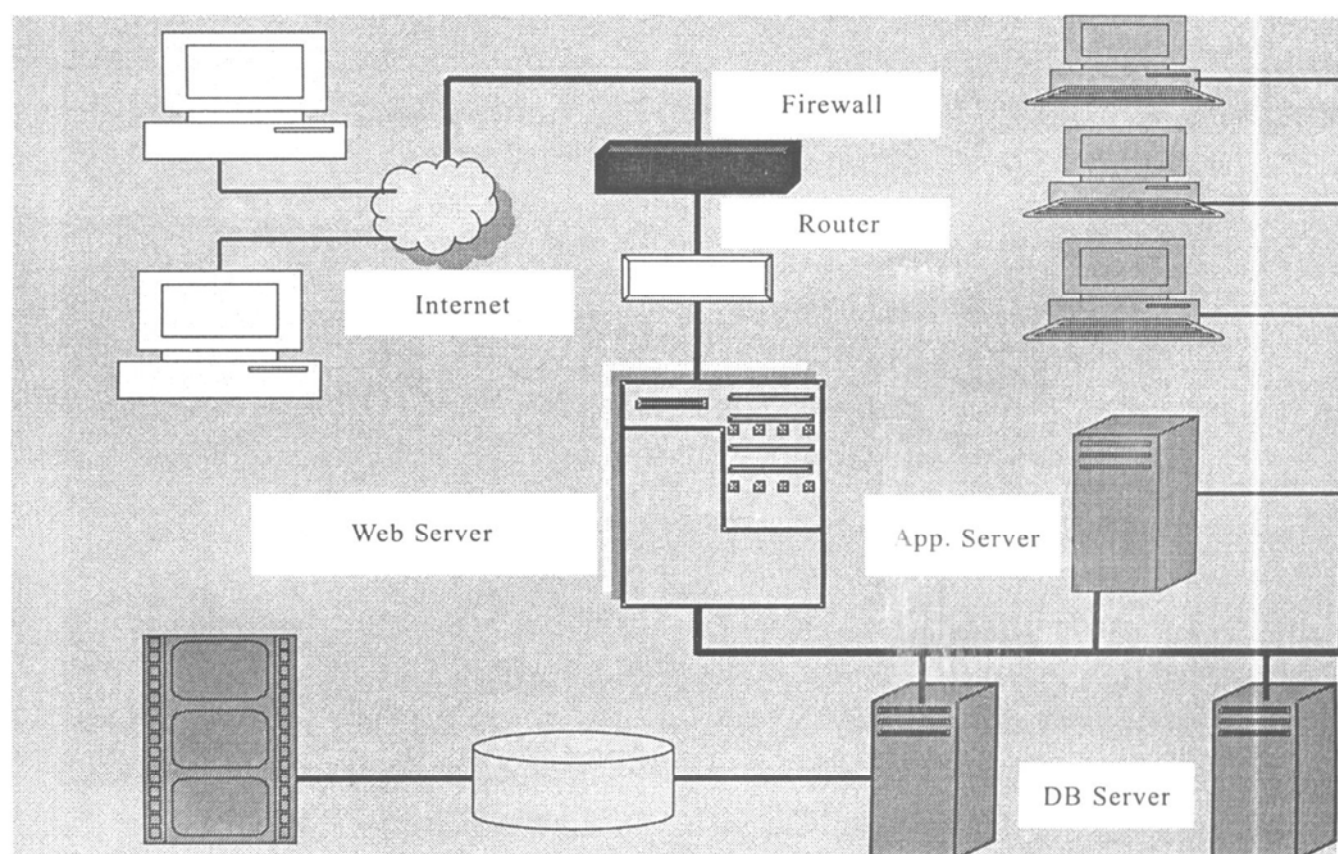


Fig 1 The Hardware Platform run in CN-PolarDB.

The software in this system includes Oracle 9i server, Internet Information Server (IIS), The link of metadata and its dataset is created in database design and development (Zheng *et al* 2002). The metadata standard (DIF) endorsed by JCADM (Cheng *et al* 2001), was chosen as the unique tool to author metadata record. More important, each data record of some dataset within CN-PolarDB, besides the dataset itself, can be searched and downloaded via Internet, whereas only data files (or datasets) can be searched and download, generally by FTP in other similar database. All data of one dataset from this system can be listed easily by browser. It doesn't require any special application software, because lots of data processing before publication is completed. But some datasets, like multimedia, remote sensing image data are stored and downloaded in database only by binary system.

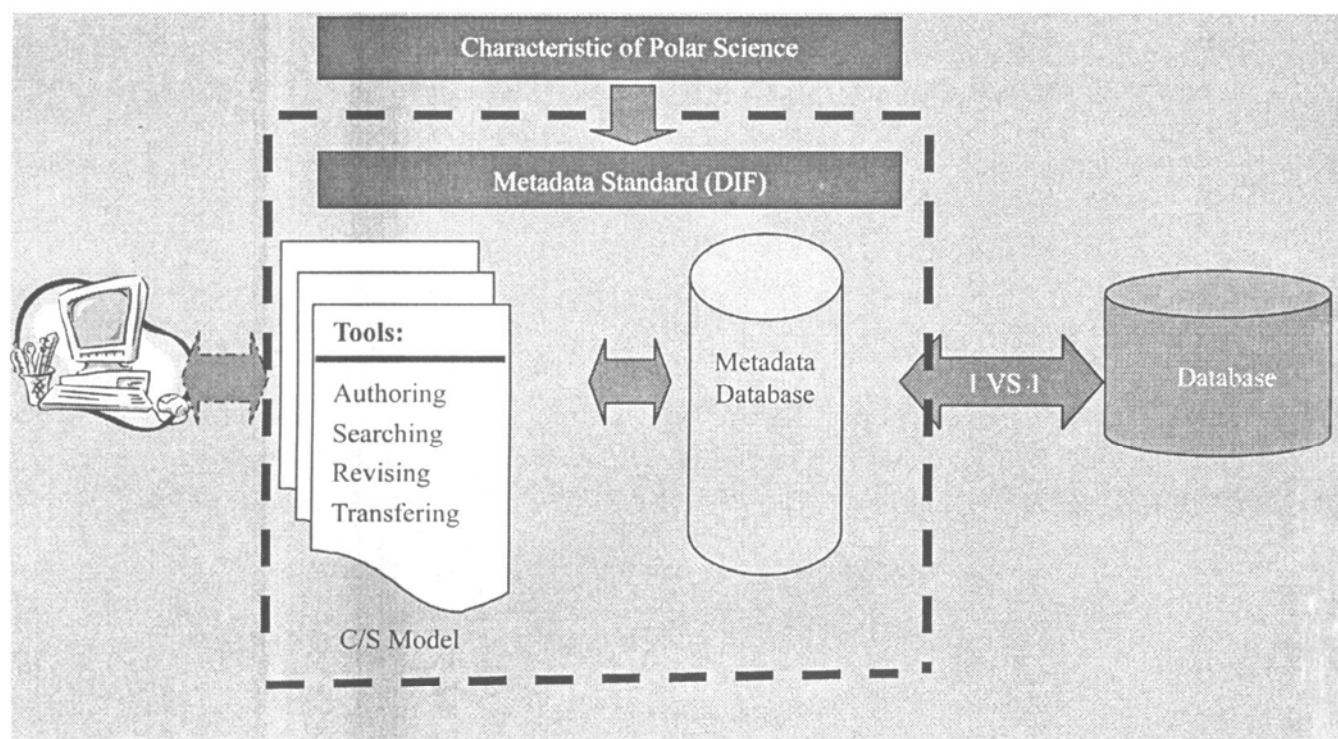


Fig 2 The process of data query.

### 2.3 The Technical Standard of CN PolarDB

A technical standard is needed for designing of a successful database, to improve the originality, accuracy of the dataset and eliminate the manmade error during the data processing. CN-PolarDB is a combined system, which involves multidisciplinary data, so it is not easy to create a feasible standard, suitable for all of these polar disciplines, as there is no similar standard to follow, unlike the metadata standard.

In the establishment of Standard of Dataset Structure, we think it is better to input all datasets into database unchanged, with each dataset independent, self-descriptive.

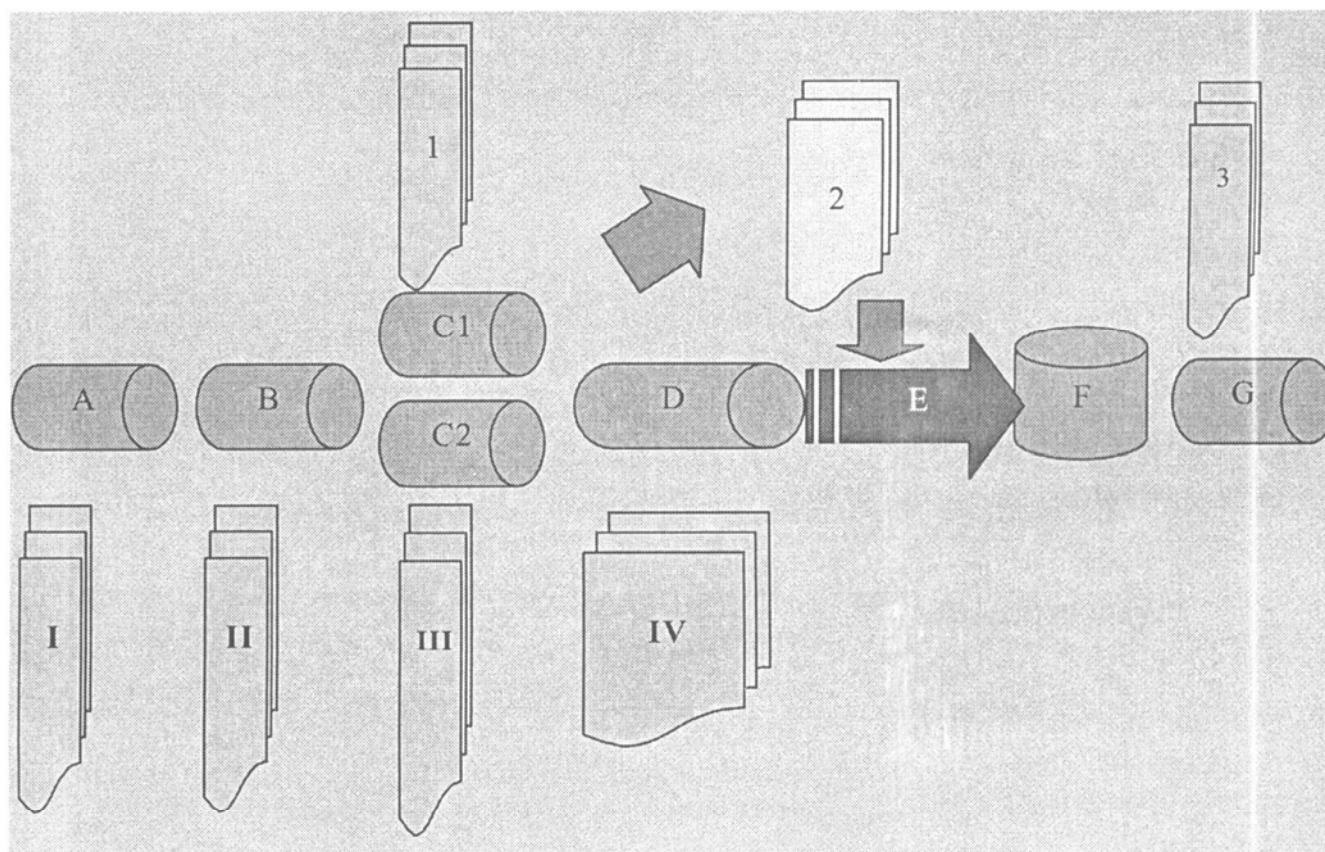


Fig 3 The flow of data processing and technical standard in CN-PolarDB

1. Standard of cataloging 2. Regulation of Quality Control 3. Regulation of Data Sharing A. Collection or Retrieving B. Registration C1. Cataloging C2. Sorting out D. Formatting E. Data Input F. Database G. Distribution or Publishing I. Regulation on Data Management II. Standard of Data Registration III. Standard of Data Preprocessing IV. Standard of Dataset Structure

## 2.4 Mechanism of Sharing Data

One knows clearly that data is of importance for science research. It is challenging to build up an easily accepted mechanism of sharing data to balance all the interests of different sides: data custodian (project leader or data collector), data coordinator, and data user.

CN-NADC is responsible for maintaining the CN-PolarDB. The data in CN-PolarDB is freely available for the scientists around the world. According to Regulation on Data Management of Chinese Arctic and Antarctic Scientific Data, published by Chinese Arctic and Antarctic Administration (CAA), its objective of CN-PolarDB development is to integrate all the data collected from polar expeditions to promote data exchange, maximize data usage, while any individual or institution has the responsibility to submit data timely (including metadata record) into CN-NADC as required.



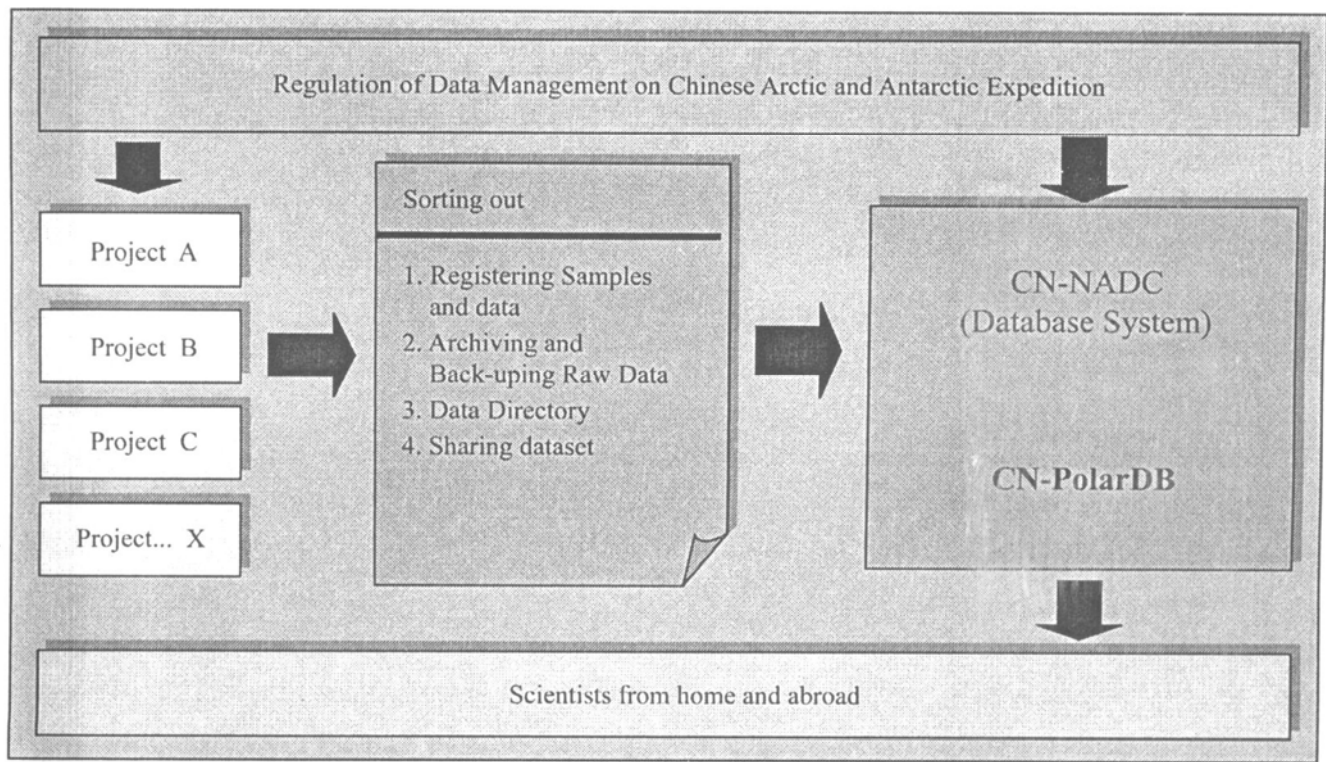


Fig 4 Framework of Sharing Data

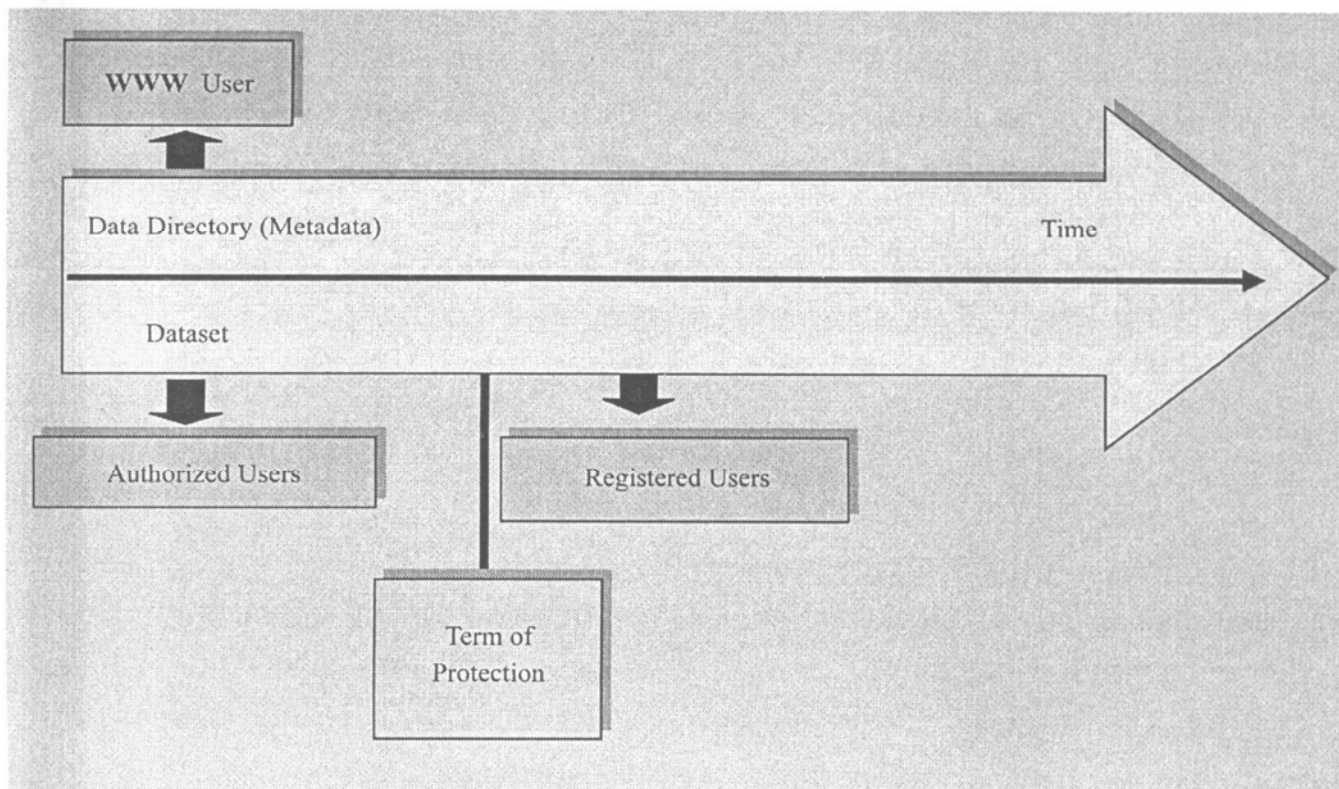


Fig 5 Mechanism of Sharing Polar Scientific data

By consensus, Manager of CN-NADC and the data custodian (project leader or data collector) define together the term of data protection, and the project leader or data collector are responsible for safe-grounding submitting data into CN-NADC commonly in two years. Within the term of protection, only authorized users have the privilege to download

datasets (however its metadata is open for the public), authorized users is generally classified by corresponding expedition voyage or key program. After expire of term of protection, the dataset is open freely for registered users. In view of it, the different user group needs to setup in this system after data is collected.

### 3 Trends of Data Management and Usage

Data acquisition, data Access and data analysis consists of three associated steps in geo-science research. With the wide-spread international cooperative research, combined with the development of database technology and Internet technology, quite a few data center, specialized in distributing the variety of dataset and data products, have setup according in which some database system have been established and their data are available via internet. Concerning polar research, the database systems in National Ice Center(USA), AAD(Australia), BAS(UK), PRIC(China) are very representative. Now, the globe data web of web-based polar databases has been formed gradually. By this data web, scientists in the world can acquire what they need if authorized. However, as the data format, quality description, and method or tool of data process differ greatly, it probably limits or reduces the data usage or effect of this data web in great extent. In this respect, the GML3( according to ISO TC211 plan, this standard will be finished before December of 2005) is expected to deal with this problem. Relatively, it become much harder to standardize the data acquisition, owing to its dependence on a range of different sensors. Draft Specification for Sensor Model Language, published by OGC working group in March of 2004, provides a general XML model for all kinds of sensor, from simple thermometer to complicated EOS. In future, researchers can easily download all kinds of data based on common format, then integrate or merge them into their own model.

CN-PolarDB has the unique feature in deal with the small, dispersed datasets, for example these datasets, even every data record of them, can be listed directly in browser, because they is transformed into Text Format before being inputted into database. In fact, the large datasets such as remote sensing data, image data, are stored at database unchanged, only can be downloaded generally through FTP. Obviously the usage of the large datasets highly depends on special software, this is not good for data integration in cross-disciplinary research. So CN-NADC and other data centers should provide lots of data products and advanced data service, to meet a range of data request, from distribution of raw data, data-reprocessing to data mining.

## References

- Zhang X, Cheng SH and Zhu JG (2002): The research and construction of the Chinese polar science database system. I The organization of data resource and research of relationship between metadata and dataset, Chinese Journal of Polar Research( Chinese Edition ), 14( 1): 2002 62- 72
- Cheng SH *et al* (2001): The research and construction of the Chinese Antarctic data directory system ( revised edition), Science Press