THE CURRENT KEY PROBLEMS AND POTENTIAL SOLUTIONS FOR GEOSCIENCES DATA SHARING IN CHINA

Youhua Ran[1]*, Xin Li[2], and Jian Wang[3]

Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China

Email:[1] ranyh@lzb.ac.cn [2] *lixin@lzb.ac.cn* [3] *wjian@lzb.ac.cn*

ABSTRACT

China's geosciences data sharing has progressed significantly but is still facing some key questions that need to be solved Because of the data's insufficiency, it has limited use in modern scientific research. The documents and metadata for the data are insufficient. The scientific data service is dated. The data application procedure is troublesome, etc. All this has become a bottleneck affecting the progress of China's scientific data sharing. Considering the reality of China, some potential solution have been proposed, which include changing the scheme of data integration, perfecting metadata and documentation, emphasizing data service, simplifying the application procedure, enhancing the shared consciousness in science and technology fields, establishing national data centers, and realizing sustainable data sharing.

Keywords: Geosciences, Data sharing, Key problems**,** Potential solutions**,** China

1 INTRODUCTION

Scientific data sharing is a political and economic necessity. It has obtained a broad consensus from the whole society. Scientific data are the most active resource for scientific and technological innovation in the information age (Sun, 2003). Geosciences activities produce geo-data, while at the same time geosciences hypothesize about and develop theory based on earth science data. More abundant, high-quality data are particularly necessary for present research on Earth systematics. Earth system science breakthroughs are frequently born in interdisciplinary frontier areas, and the interdisciplinary and multidisciplinary fields need support from multidisciplinary knowledge, information, and data.

The development of science and technology highlights the strategic importance of sharing scientific data. The lack of high-quality and systematic geo-scientific data has become a great bottleneck in scientific and technological innovation as earth system science advances into a new era. The management and orderly sharing of mass amounts of geo-science data will greatly promote the continual development of Earth system science research and provide powerful support for long-term technology development and breakthroughs. At the same time, scientific data are society's shared knowledge resources; they should be shared to serve the state's overall goals (Cheng et al, 2004). On the one hand, massive scientific data are not presently being fully utilized; on the other hand, many needs of scientific data are not met. The obvious contradiction between the supply and demand of scientific data has become a pressing challenge.

In recent years, Earth science data sharing in China has made tremendous progress by the combined efforts of many organizations. By the funding of the Ministry of Science and Technology in China, the Chinese Academy of Sciences, the China Meteorological Administration, the China Ministry of Water Resources, the National Natural Science Fund Committee, etc., an open, orderly scientific data sharing environment has gradually been built. The key basic scientific data have been gradually shared by scientists and society and a number of scientific data sharing platforms have opened, for example:

- The China meteorological data service system (http://cdc.cma.gov.cn),
- The National MODIS data center (http://satellite.cma.gov.cn/eos/),
- The data-sharing network of Earth systematic science (http://www.geodata.cn),
- The Resource and Environment Data Center of the Chinese Academy of Sciences(http://www.resdc.cn),
- The scientific database of the Chinese Academy of Sciences(http://www.sdb.ac.cn),
- The Chinese Terrestrial Ecosystem Flux Research Network (ChinaFLUX) (http://www.chinaflux.org),

- The national satellite weather center (NSMC), and
- The satellite climate share data network (http://scsd.cma.gov.cn).

The project of the Environmental and Ecological Science Data Center for West China (WESTDC) (http://westdc.westgis.ac.cn) was supported by the National Natural Science Foundation of China's (NSFC) "Environment and Ecological Science Project" in 2005. The project is undertaken by the Cold and Arid Regions Environmental and Engineering Research Institute (CAREERI) and the Institute of Geographic Sciences and Natural Resources Research (IGSNRR). Earlier, the two major institutes in resources and environmental studies areas had made a long-term exploration into scientific data sharing in China. They had many pioneering works and accumulated rich data development, management, and scientific data services experience. From April to August 2006, the WESTDC data services group visited the project principally supported by the National Natural Science Foundation of China's (NSFC) "Environment and Ecological science Project." They have communicated extensively with the project principal to give them a first-hand profound understand of our problems in scientific data sharing from different users' perspectives. We found these problems are common in China. By considering this reality as well as the possible options, this paper tries to give some potential solutions to these problems and to illuminate the future for WESTDC building and geo-science data sharing breakthroughs in China.

2 STATUS AND KEY PROBLEMS

In recent years, researchers have done much work on Chinese geo-science data sharing. Many platforms for data sharing have opened and the corresponding services are currently in operation. Various data such as remote sensing images, meteorology, hydrology, ecology, soil, desertification, basic geographic mapping data cryosphere, and other basic data can be obtained through online or off-line means. That the data environment is being greatly improved also contributes to the development of Chinese geo-scientific research.

After discovering the conditions of existing geo-science data sharing platforms and the implementation of the data sharing engineering and through analyzing the data and the data services provided by existing data centers, we must understand that current Chinese scientific data sharing still has many deficiencies both for the data user and for the platform builders. The development of the career of geosciences data sharing still has many systematic and institutional obstacles. There is an urgent need to resolve many key issues.

Firstly, the data for sharing are very limited. Although much data can be downloaded online or applied offline, most data are limited to satellite data and large-scale background data. Although huge amounts of data exist, the data categories are incomplete, and the lack of systematic data has little value for most research;

Secondly, the data standards are inconsistent; metadata are missing, and it is hard to evaluate data quality. Data provided by most platforms are not detailed documents, and metadata are very limited. It is difficult for data users to evaluate the research results and make reference to them, which affects the data authors' enthusiasm for using these services.

Thirdly, the service consciousness of most data managers is absent and inefficient, and the data application procedure is tedious. Most data centers provide a single type of data service, while users get the data only after filling out paper applications, getting signatures and stamps, and other cumbersome processes. Furthermore, the sharing of scientific data often takes a long time with many man-made barriers restricting the process.

Fourthly, there are few people who do data sharing. Today scientific data sharing projects are mostly carried out by researchers. In academia, scientific data management, development, and service are not considered to be a science (Sun et al, 2003). Many people do the work in a passive mood. Meanwhile, there is no good assessment to improve the staff's enthusiasm, which leads to incomplete data sharing and poor data service quality.

Fifthly, the accessibility of many data sharing platforms is poor. Data sharing platforms are built for specific projects. After a project is finished, the platform operating cost usually can not be guaranteed. Therefore this has a negative effect on the sustainable platform operation and on the confidence of users.

3 POTENTIAL SOLUTIONS

After analyzing the problems in Chinese scientific data sharing in the above paragraphs, we see that the problems interact with one another. We will discuss the potential solution for the above problems from the point of view of the

data center builders and the government based on reality and possibility.

3.1 Change the scheme of data integration to enrich the database

The scarcity of data resources is one of the most crucial issues in current geo-scientific data sharing. Currently, we collect data depending on the financial support of data projects and limited system support. The data collected are very limited, and most of them are part of background data in geosciences research. Data quality is difficult to guarantee. Certainly, at the beginning of data sharing, this scheme of collection has played an important role. After the background data collection is finished, collecting data has small potential and is difficult. In order to maintain the sustainable development of data centers and maintain an adequate data supply, the existing data collection methods must be innovated.

Scientific data are characterized by the special production research, and data output is flexible and changeable. These groups interested in scientific data are complex. We must consider addressing the problem of multiple-interests and do long-term planning to collect the needed data. According to international experience, some suggestions are as follows:

- Find data through literature. This will ensure data quality and also greatly reduce the difficulty of data collection because data authors would like to share in order to enhance the rate of article citations.
- Cooperation with other data sectors. There is great progress in basic data sharing although not all researchers can use these data directly. The reason for this is that data sharing in China is spread out and the work is distributed to several sectors separately, which necessitates that data users fill out multiple data applications for each data center to meet their data needs. Therefore, strengthening the cooperation among basic data departments to provide one-stop data service for users is very important. It is time to establish a national data center by integrating the existing data centers.
- Actively supporting data mining for existing data and strengthening cooperation with other departments in the data area to continue rescuing historical data. It is necessary to mine the existing data in hot fields. For example, to complete more high-precision data for several large inland river basins, it is important to produce and collect biophysical parameters to model data sets of these areas and to collect major national projects such as the Qinghai-Tibet Railway zone data.

3.2 To perfect metadata and data documentation and establish data sharing standards and achieve standardization of data management

Metadata are the data that explains data. For geospatial data, such metadata information includes data abstract, format, projection, downloading, and reference. Data documents are records that data producers make during the course of data production, including important information such as data production methods, thinking, and so on. We found that the above-mentioned information is invaluable to unanticipated data users (Parsons et al, 2005). Moreover, detailed metadata information and data documentation are improving data access rate.

Standardized data not only helps the user, it is also helpful for data management. This standardization includes the data format, projection, etc. Therefore, data standardization is very important for both the data manager and data exchanger. First, however, data standardization needs to make certain data standards are reasonable. For these data standards, we should consider the corresponding international standards. Building national data standards for the sharing of scientific data, including standards of data storage and distribution, is an important foundation for improving the efficiency of scientific data service and database integration. However, standardizing metadata, data documentation, and data standards is challenging. At present, we should pay equal attention to file documents such as data in current data production, management, and sharing. We suggest that the personnel working with needed documentation should come from experts in the corresponding field.

3.3 To emphasize data service and simplify application procedure

Data are the basis of data sharing; data service is the core of data sharing. Data services should be scientific, and their content as well as form should be diversified. Providing reliable data is the most basic element of data services. Scientific research work has a traditional openness, and people engaged in scientific data sharing are asked for a more active and open attitude. Only in this way will the unnecessary restrictions be removed and the scope of data

service be expanded to provide the greatest degree of data support for geosciences research. The major reason for the inefficient and complex data application procedures for the science data services in China is the lack of functional improvement and advanced technology in the scientific data sharing platform. Therefore, we should strengthen our geosciences data sharing talents, particularly platform construction talents, making full use of the advantages of the Internet to simplify data application procedures and to enhance the efficiency and quality of service.

3.4 To enhance the sharing awareness in the science and technology community and start a data-sharing system building

Currently, the lack of a support system for scientific data-sharing and the need for a system for its long-term operation is one of the greatest strategic issues in scientific research. The Chinese scientific data sharing policy environment has not been established, data property is confused (Liu, 2003), and the sharing awareness of the science community is not high. For example, WESTDC has received many telephone calls from data users for advice on data applications who remark, "So handy?" after learning about the application procedure. We found that potential users' expectations are low when encountering a long blockade of all data for data sharing. This requires that data managers and data authors change their ideas, putting the national interest first, so that departments and individual interests serve the national need and actively work to contribute to the sharing of scientific data. Data sharing staff should actively explore scientific data sharing - driven mechanisms to enhance the overall understanding of the scientific community about sharing data, which will greatly promote the development of China's progress in geosciences sharing and data sharing system formation. It will also be conducive to the effective implementation of the data sharing system. The data sharing system should be formulated to protect the rights of investors and data authors on the basis of state secrets. It must also balance the interests of all parties, as based on the government-driven policies. When the time is ripe, the corresponding laws and regulations to standardize the action of scientific data sharing will be developed.

3.5 To establish national data centers and realize sustainable data sharing

After experiencing decades of free development, the basic pattern of Chinese geosciences data sharing has been formed. For building national data centers, the government should correct long-term special funds and establish the corresponding laws to support their operation (Chen, 2003) to lead to the formation of an orderly data sharing environment. For this approach, we can refer to a proper management scheme of scientific literature to scientific data publishing and sharing.

4 CONCLUSION AND DISCUSSION

Good scientific data stewardship requires explicit recognition and understanding of data user communities (Parsons et al, 2005). We have much to learn from international data communities, such as the United States Snow-ice Data Center (NSIDC), the NOAA Climate Data Center (NCDC), and the Land Processes (LP) Distributed Active Happening Center (DAAC), etc. However, we must understand that the contradictions between the supply and demand of scientific data will exist forever. The development of geosciences data sharing still has many institutional obstacles. The solutions proposed are synergistic and can act on the data sharing through a common framework.

For scientific data center builders, the first problem is to solve data integration and strengthen cooperation and exchange with other data centers to maintain scientific data sharing harmony and sustainable development. They must cultivate scientific data work talent, especially scientific data platform construction teams, promote science data sharing standards, and achieve standardization of data management. They also need to improve metadata and data documentation, strengthen the work and consciousness of data services, and improve the efficiency of data services. The government should employ specific measures to promote the entire scientific community's understanding of the problem of data sharing and improve data sharing awareness. Moreover, establishing national data centers is basic to promoting data-sharing system building and sustainable scientific data sharing.

5 ACKNOWLEDGEMENTS

This work was supported by the National Scientific Foundation of China (90502010 and 40401044). The authors also thank the anonymous reviewers for their extremely helpful and constructive comments on the manuscript.

6 REFERENCES

Cheng, J. (2004) Data sharing—A Strategical Choice Facing the Scientific Community (in Chinese). *China Basic Science*, pp.1-5.

Chen, j. & Wang, C. (2003) Consideration on Mechanism of Scientific Data Sharing (in Chinese), *China Basic Science*, pp.40-43.

Duerr R., Parsons, M.A., Marquis, M., Dichtl, R., & Mullins, T. (2004) Challenges in long-term data stewardship. *Proc. 21st IEEE Conference on Mass Storage Systems and Technologies.* NASA/CP-2004-212750 (pp.47-670).

Fetterer, F. & Knowles, K. (2002, updated 2004) *Sea Ice Index*. Boulder, CO: National Snow and Ice Data Center. Digital media. Retrieved 13 April 2005 from the National Snow and Ice Data Center Web site: http://nsidc.org/data/seaice_index/.

Liu, C. (2003) Management of National Scientific Data Sharing in USA and Its Enlightenment on China (in Chinese), *China Basic Science*, pp.34-39.

Lu, Y. (2000) Sharing Space Data and Its Ethical Challenges, *Journal of Remote Sensing 4(4)*, pp.245-250.

Parsons, M. & Duerr, R. (2005) Designating user communities for scientific data: challenges and solutions. *Data Science Journal 4(0)*. pp. 31-38.

Sun, J. (2003) Data sharing needs the legal support. Retrieved 17 Feb. 2005 from the Daily online of the Chinese People's Liberation Army website: http://www.chinamil.com.cn/gb/pladaily/2003/02/17

Sun, S. (2003) Geo-data is the important sources for the innovation of geosciences—discusses the Geosciences data sharing for Geosciences (in Chinese). *Advance in Earth Sciences 18(3)*, pp.334-337.