

STUDY ON DATA INTEGRATION AND SHARING STANDARD AND SPECIFICATION SYSTEM FOR EARTH SYSTEM SCIENCE

Wang Juanle, Sun Jiulin

Institute of Geographic Sciences and Natural Resources Research, State Key Lab of Resources and Environment Information System, Chinese Academy of Sciences, 100101, Beijing, China - (wangjl, sunjl)@igsrr.ac.cn

KEY WORDS: Earth system science, Geosciences data sharing, Standards and specifications, Geosciences Facility

ABSTRACT:

Earth System Science (ESS) has evolved into a new historic stage beyond Earth Science. ESS takes the whole earth systems and their interactive actions among spheres as its objective. This causes that its development need lots of multi-disciplinary, multi-sources, multi-type, integrated geosciences data resources support. According to this requirement, Data Sharing Network of Earth System Science (DSNESS) was established at the beginning of China Scientific Data Sharing Program (SDSP) launched in 2002. Data sharing in DSNESS need standards and specifications environment. Basic data sharing concept model are studied firstly for the requirement. According to 4 principles designed, ESS data sharing standards and specifications system and its relationship among related international and domestic standards system are studied. Designed standards and specifications system includes 4 main classes, i.e., mechanism and rules class, data management standards and specification class, platform development specification class, data service specification class. Total 18 standards, rules and specifications have been drawn up. Through almost 6 years research and application, nowadays all of these 18 standards and specifications have been used successfully in the distributed data sharing network of earth system science which include 1 general center and 13 sub-centers. In the near future, standards and specification environment of DSNESS will be further development two directions oriented, one direction is basic concept reference model research, and another direction is data fusion and assimilation standards and specifications study.

1. INTRODUCTION

Geosciences have evolved into Earth System Science stage in 21st century (Liu Dongsheng, 2004). Because Earth System Science (ESS) takes the whole earth systems and their interactive actions among spheres as its objective, it has series typical features, such as global system, whole temporal and spatial scale, multi-disciplinary integration, high technology application, global collaboration, and so on. Usually, its research approach includes observation, understanding, simulation and global environment change prediction (Zhou Xiuji, 2004. Huang Dingcheng, etc., 2005). Its development need lots of multi-disciplinary, multi-sources, multi-type, integrated geosciences data resources support according to its disciplinary features and characteristics. While, most of these geosciences data are distributed in different research agencies, group or even personal scientists' database. How to integrate and share these data in Earth System Science research community is a challenge for ESS development.

According to this requirement, China has begun to take actions for scientific data sharing since 1994. Through about 10 years' preparation, China Scientific Data Sharing Project (SDSP) was launched in 2002 (Xu Guanhua, 2003). Data Sharing Network of Earth System Science (DSNESS) is one of the first 9 pilot projects in SDSP (Sun Jiulin, 2003). It has been one of long term data sharing platforms in National Science & Technology Infrastructure (NSTI) since 2005. This project will provide Earth System Science data sharing platform for scientists. Its objective is to enhance and support the Earth System Science research and science & technology innovation through integration and sharing of distributed scientific data in institutes, universities, individuals, government funded research projects, and international organization or data centers. At present, DSNESS' data resources mainly focus on 'Land surface and human-land relationship' research Area.

2. STANDARD AND SPECIFICATION REQUIREMENTS

Most of the Earth System Science data are produced in the processing of science and technology research activities, especially from the science and technology research projects funding by government. These data are different with the data derived from the long term operational system by varies agencies or departments, such as meteorology observation data, earthquake monitoring data and so on.

National Science Foundation of USA make a clear definition for research data and other data in its report, "National science board: Long-lived digital data collections: enabling research and education in the 21st century" published in sep. 2005. In this report, data collections are divided into 3 types, i.e., research data collections, resource or community data collections, reference data collections. Research data collections are the products of one or more focused research projects and typically contain data that are subject to limited processing or curation. They may or may not conform to community standards, such as standards for file formats, metadata structure, and content access policies. Quite often, applicable standards may be nonexistent or rudimentary because the data types are novel and the size of the user community small. Research collections may vary greatly in size but are intended to serve a specific group, often limited to immediate participants. There may be no intention to preserve the collection beyond the end of a project.

Resource or community data collections serve a single science or engineering community usually. These digital collections often establish community-level standards either by selecting from among preexisting standards or by bringing the

community together to develop new standards where they are absent or inadequate.

According to the data collection definition of NSF, ESS data belong to typical research data. Integration and sharing of ESS research data need standard and specification environment.

3. STANDARD AND SPECIFICATION CONCEPT MODEL

Earth System Science data sharing standard and specification environment concept model is shown in figure 1. As left side of figure 1 showed, data derived from “project A” is only used in small user group inner project A. These data will not be shared for other users. This make these data is difficult to be found and accessed. And even some users find these data luckily, they can’t understand and use them easily, because these data have different catalogue classification, content structure, concept semantic, data format etc. Similar with this, project B’s data will only be used by the scientists in project B. other users can’t acquire and use them without the data sharing environment. This leads large number of geosciences data can’t be used by more scientists, and geosciences data can’t play important role for the geosciences development.

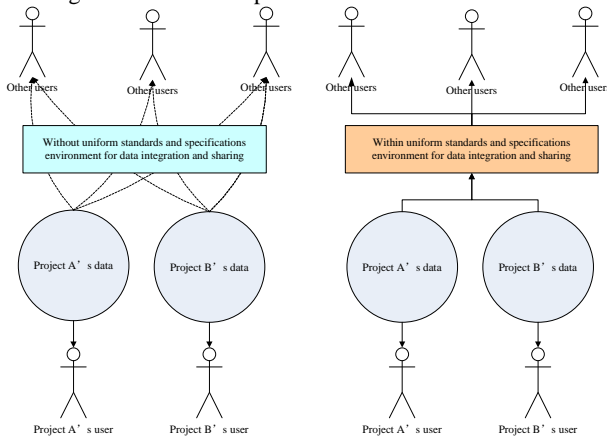


Figure 1. Geosciences data sharing environment

As shown in right side of figure 1, if we establish a data sharing standard and specification environment for Earth System Science, project A and project B’s data can be searched, browsed, accessed and acquired with uniform standards for more scientists.

But it is not difficult to establish this environment. Concretely, there are 3 main problems needed to be studied at present, i.e., how to establish the ESS data sharing mechanism? How to integrate and share multi-disciplinary data? How to make data easily for user find and access?

4. BASIC PRINCIPLES FOR DATA SHARING STANDARD AND SPECIFICATION ENVIRONMENT

Through analysis, there are 4 basic principles should be paid more attention to for Earth System Science data sharing standard and specification environment construction.

4.1 Corresponding principle

Standard and specification should correspond with their inside system and outside systems. First of all, all the standards and specifications should avoid conflict inside the Earth System Science data sharing system. At the same time, standard and specification of Earth System Science should correspond with its out side system. For example, Earth System Science data sharing is one of important components of NSTI in China, so it should keep consistency with the standards and specifications in NSTI system. Also, these standard and specification should keep consistency with the international standards, such as ISO 19115 metadata standard, etc.

4.2 Easy access principle

The final objective of Earth System Science data sharing is to provide an easy collection, management and dissemination environment. Let the user easily access all the Earth System Science data is a basic principle for the whole standard and specification system. Under this principle, Earth System Science data should show perfect and easy understand data catalogue for users firstly.

4.3 Reference model instruction principle

Geosciences data have a complex process from the data collection to dissemination. There should be a reference model to instruct the data sharing standard and specification environment. For example, ISO geography information reference model is a good example for Earth System Science standard and specification environment.

4.4 Software implementation principle

Data sharing need information and technology support. At present, distributed network is a popular tool for data sharing. Based on Earth System Science standard and specification concept model, its logical and physical model should also be researched. This principle will contribute to the software developers develop data collection and management tools, which is a good method for the standard and specification implementation in science community.

Based on the basic principle 1 and 2, Earth System Science data sharing standards and specifications should keep consistency with national and international standard systems. As figure 2 shows, Earth System Science standard and specification system should reference the ISO 19100 geography information series standards, Open Geospatial Consortium (OGC) interoperability technology standards, China national basic geography information standards system, China e-government data sharing standards system, NSTI basic standards system, SDSP basic standards system, etc.

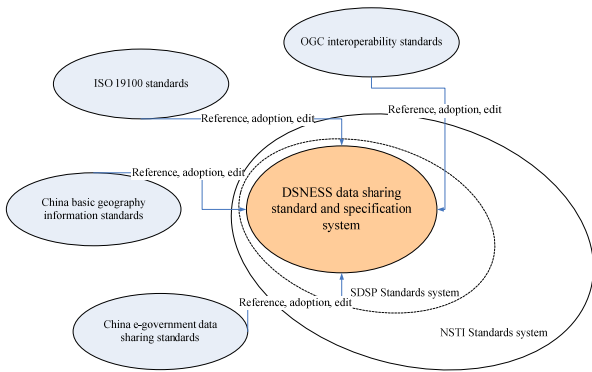


Figure 2. Orientation of standard and specification system for Earth System Science data sharing

5. SYSTEM ARCHITECTURE

Based on the principles and orientation of standards and specifications system, DSNESS established this initial architecture through almost 6 years of study and practice. The general architecture includes 4 class standards and specifications (shown in figure 3). They are mechanism and rules class, data management standards and specification class, platform development specification class, data service specification class.

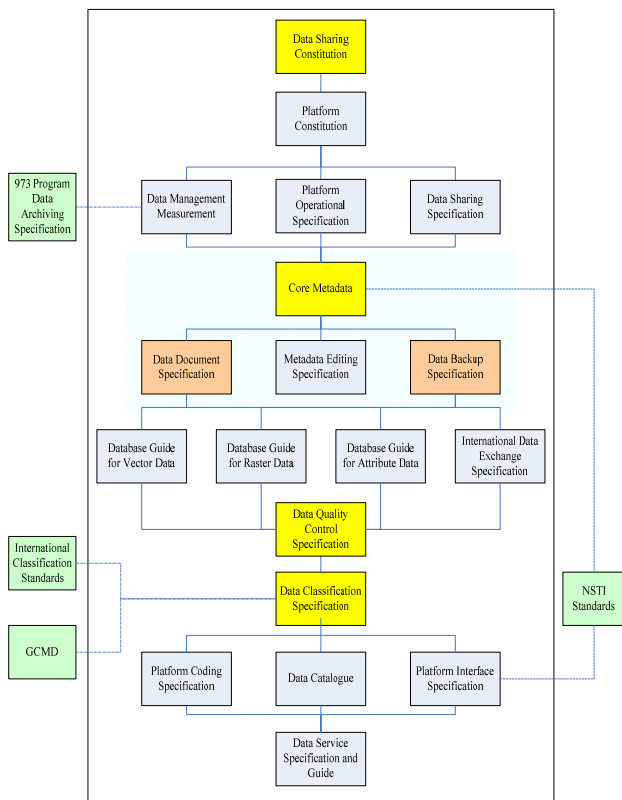


Figure 3. Standard and specification system structure for Earth System Science data sharing

5.1 Mechanism and rules class

ESS standard and specification mechanism and rules class include data sharing Constitutions, platform implementation measurement, platform operational management rules, data sharing rules and guides. These rules keep consistency with the

national data sharing law and rules. For example, it will under the law of “Science and Technology Progress Law of China”, “Specification of Scientific Research Program data archiving in Resources and Environment fields of National Key Basic Research Program” (Wang Juanle, etc., 2009), etc.

Data sharing constitution is the core mechanism of DSNESS. This constitution’s core idea is to call for those geosciences research institutes, university, data centers or organizations, and personal scientists joining into this union. Data sharing organization architecture, union members’ rights and responsibilities and duties are designed in this union. There are about 20 fixed members in this union at the beginning stage.

5.2 Data management standard and specification class

Data management standard and specification class includes metadata standard, metadata editing specification, data document specification, data backup specification, international data collection and exchange specification, data quality control specification, database design specification for vector, raster and attribute data, etc. It also keeps consistency with the national specifications. For example, DSNESS metadata standard is coordinated with NSTI core metadata standard and ISO 19115 geography information standard.

In these specifications of data management, metadata standard is the core one for data integration and sharing standards. Metadata, known as “data about data”, describes data content and related dissemination information. Geosciences metadata, usually presented in XML format, represents who, what, when, where, why and how about a piece of geosciences data or other resource. Metadata possesses many benefits in data archive, assessment, management, discovery, transfer, and distribution. In the geosciences data context, metadata has the following usages: discovery of resources, Evaluation of resources, Use of resources, Contract between the user and the provider.

There are object level metadata and collection level metadata in DSNESS. Object level metadata describes a single entity. Collection level metadata describes a series or a group of entities. Geosciences metadata typically includes the following:

- Identification: including dataset title, citation, abstract, purposes, and keywords, etc.
- Quality: including positional accuracy, data completeness and consistency, etc.
- Spatial reference: including coordination system, spatial extent and temporal coverage of the data, etc.
- Distribution information: including distributor of and options for obtaining the data set, such as the format of the resource, the URL to download the data or access the Web services, etc.

Core metadata standard of DSNESS include 188 metadata elements, including 22 core elements.

5.3 Platform development specification class

Platform development specification class includes data classification specification, software development and coding specification, software interface specification, etc. Earth System Science data sharing classification specification is the core specification in this system. DSNESS designed the time scale classification, spatial scale classification, data catalogue classification and related coding system. This classification is the basic for data management and access. It is also the basic

for the platform software development. According to the basic principles, DSNESS data classification specification is consistency with china's disciplinary classification and code standard (GB/T13745-92) and Global Change Master Directory.

5.4 User service specification class

It includes user service specification and data service guides. These specifications ensure perfect user services for data sharing.

6. IMPLEMENTATION AND APPLICATION

At present, about 18 standards and specifications in this architecture have been drawn up. All of them have been used in DSNESS till now. Based on the organization of DSNESS, one general center and 13 distributed sub-centers all use these specifications and standards.

Under this standard and specification environment, more than 18TB data resources have been collected and shared in this network. Till to the end of 2009, almost 46000 users have resisted in the platform, 24.58 TB data have been downloaded for public. Figure 4 shows the website homepage of DSNESS (<http://www.geodata.cn>).



Figure 4. Website home page of DSNESS

7. DISCUSSION

Earth System Science data sharing need a long term processing and its standard and specification environment should be revised and fulfilled continually. Face to the near future, there are two directions that these standards and specifications should be developed urgently. One is DOWN direction and another is UP direction. Under the DOWN direction, the basic concept model of standard and specification environment should be researched and studied deeply. Under the UP direction, different stage's concrete standards and specifications according to the data sharing life cycle should be paid more attention to, especially for the data product specifications, dataset fusion and assimilation specifications, dataset service specifications, etc.

8. REFERENCES

Liu Dongsheng, 2002. Global change and sustainable development science. *Earth Science Frontiers*, 9(1), pp. 1-9.

Zhou Xiuji, 2004. Some cognitions on earth system science. *Advance in Earth Science*, 19(4), pp.513-515.

Huang Dingcheng, Lin Hai, Zhang Zhiqiang, 2005. *Strategy Study on Earth System Science Development*. Beijing: Meteorology Publishing House.

Xu guanhua, 2003. Advance for enhance China's science and technology innovation capacity by data sharing, *China Basic Science Research*, (1), pp. 5-9.

Sun Jiulin, Shi Huizhong, 2003. Construction of Earth System Science Data Sharing Network in China. *China Basic Science Research*, (1). pp. 76-82.

Sun Jiulin, Lin Hai, 2009. *Earth System Study and Scientific Data*. Beijing: Science Press.

National Science Foundation. Long-lived Digital Data Collections: Enabling Research and Education in 21st Century , September 2005

Wang Juanle, Sun Jiulin, 2007. Development of China WDC Systems for Data Sharing, *China Basic Science Research*, (2), pp. 36-40.

Wang Juanle, Yang Yaping, Zhu Yunqiang, etc. 2009. Data Archiving Progress and Data Types Analysis of National Basic Research Program of China (973 Program) on Resource and Environment Field. *Advance in Earth Science*, 24(8), pp. 947-953.

NSF Board, 2005. Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century, <http://www.nsf.gov/pubs/2005/nsb0540>

9. ACKNOWLEDGEMENTS

The research was funded by MOST (2005DKA32300); NSFC (40801180); Open Research Funding Program of LGISEM; Forward project for youth researcher of IGSNRR; Informationization Project of CAS. Also thanks Prof. Teng Mianzhen, You Songcai, Chen Shenbin and other colleagues' contribution in this research.