Scientific data access activities and infrastructure in China

Jianhui Li[1,2], Wangping Li[1], Baoping Yan[1], Hui Zhang[1,2]
1 Computer Network Information Center, Chinese Academy of Sciences,
Beijing 100190 China
2 Chinese National Committee for CODATA, Beijing 100190 China
Email: lijh@cnic.cn

Science research has become increasingly data-intensive, scientific data has become the engine of discovery. Great efforts are being made in China to promote science and technology data and information sharing. This paper introduced several national level scientific data sharing activities, including the Scientific Data Sharing Program(SDSP), and National Science and Technology Infrastructure Platform(NSTI). CAS as highest academic institution in natural sciences in China is also promoting to build data infrastructure , not only providing service for CAS innovation but also open accessed by others . The scientific data infrastructure (SDI) provides a transformed scientific method and power science innovation. So the scientific data infrastructure is the key component of Cyberinfratructure and e-Science in CAS. This paper gave the framework of scientific data infrastructure in CAS and introduced the progress of construction and service in detail. The Scientific data infrastructure in CAS includes massive data storage system, large scale research databases, data integration middleware and data analysis tools, and application enabled environment. Based on the massive data storage system and cloud storage service, which linked the supercomputing facilities by CSTNET（China Science and Technology Network）, the Scientific Databases（SDB）, one large scale, distributed, multi-disciplinary research databases system, can be open accessed. In order to integrate all these databases, the scientific data grid middleware has been developed. The application enabled environment, which not only Integrating scientific data, storage, computing, analysis model and tools, but also providing an easily and friendly interactive interface for users, is one essential part of the scientific data infrastructure. Several typical pilot applications based on this environment be discussed in detail.