Bilateral US-China CODATA Workshop 2014

Evolution of Chinese Research Data Policy

Jianhui Li (lijh@cnic.cn)
Computer Network Information Center, CAS
CODATA-China
25 Aug 2014
Outline

• Scientific Data Sharing Program (SDSP) in China
• Open access Research Information of NSFC
• Scientific Data sharing in CAS
• Conclusion
Scientific Data Sharing Program in China

• Promoted by Scientist, Ministry of Science and Technology (MOST) gives top priority to scientific data sharing and includes the implementation of Scientific Data Sharing Program (SDSP) into the construction of national science and technology infrastructure platform
  - a national level program was launched by MOST in 2002.
  - **Pilot project, 2002-2005**
  - **Phase I**: 2006-2010, infrastructure construction and data integration
  - **Phase II**: 2010-, open access service, evaluation and authority
Goal Of SDSP in Phase I

- By the end of 2010, build a data management and sharing service system with a three-tier structure of 40 scientific data centers or networks covering the 6 disciplines of natural science and environment, agriculture, population and health, basic and frontier sciences, engineering and technology and comprehensive regional science.
SDSP activities in Phase I

• More than 18 scientific data sharing projects were funded by MOST after 2002.
  – Covering Resource and Environment, Agriculture, Population and Health, Basic and Frontier Sciences, Engineering and Technology, Regional Development fields in 24 departments.
  – Establishing the data sharing environment, including policy, standards, data products, data sharing platform, etc.
  – Promoting the data sharing for all society
Research Data Sharing Policies

• National level
  – Promoting to Add scientific data sharing section to China Scientific and Technological Progress Law

• Organization Level
  – 40 rules and regulations on data sharing made by relevant departments in various domains
  – For example, Meteorology Data Sharing Management Specification Published by CMA. National Key Basic Research Program (973) Data Archiving Specification on Resources and Environment fields.
Clearinghouse and Service System

• Open Access Portal provided service in September, 2009 – China Science and Technology Resource Sharing Network

• A national level research data clearinghouse be opened to public
Research Data Catalogue

• population and Health
  – Bio medical, clinical, public health, traditional Chinese medicine, pharmacy, population and reproductive health, ect.

• Earth system
  – Geographical, Natural Resources, Ecological and Environment, Polar Research, Space Science, earth Observation, ect.

• Forestry
  – Forest resource, forest protection, forest cultivation, wood science

• Meteorological and atmospheric science
• Agriculture Science
• Earthquake Science
• Basic Frontier Science
  – Physics, chemistry, astronomy, material science, biology, ect.
Activities in Phase II

• Six research data system be authorized as national science and technology infrastructure
  – Forestry Science Data System
  – Data Sharing Network of Earth System Science
  – Data Sharing Network of Population and Health
  – Agriculture Science Data system
  – Meteorological Science Data System
  – Earthquake Science data system

• Make service assessment annually
  – the volume of research data can be open accessed, users and user case, capacity, sustainable, etc.
What we learned

• Pilot-projects can driven research data sharing in the start stage
• Polices and mechanisms are more important than funding
• We need make clear policy on research data sharing in national level
• Related regulations and laws should be developed
Open access Research Information of NSFC

- The National Natural Science Foundation of China (NSFC) is one of the biggest funding agency for research in China.
- 35000+ new awards each year with 17-20 billion.
- Huge “long-tail” research data produced by these awards.
Open access Research Information of NSFC

- A specific web site opened on 2006
- From 2012, opened projects information and related achievements, for example papers, books, etc.
- Now including 92585 projects, and related 1367678 items
What we learned

• NSFC began to open grants information, and have clear regulation to open research results, especially for papers and books

• NSFC should have clear policy for research data sharing produced by their awards
Scientific Data sharing in CAS
Scientific Data Deluge in CAS

• Large scientific facilities produce huge data
  – +20 being operation
  – +20 under construction

• Long-Term field observation stations
  – +100 stations including Ecology, Environment, Space, etc.

• Long-Term Research data need to be archived and curation and sharing
  – 100+ institutes
Scientific Databases (SDB)

• A Long-term mission started in 1986 which funded by CAS
  – many institutes involved
  – long-term, large-scale collaboration
  – data from research, for research

• Collecting multi-discipline research data and promoting data sharing
  – More than **350** research databases and **500** datasets by 61 institutes
  – Over **200TB** data available to open access and download

http://www.csdb.cn
Scientific Databases (cont.)

- focusing on data integration and improving research database to be resource database and even reference database)
Scientific Databases (cont.)

- 8 Resource databases
  - Geo-Science
  - Biodiversity
  - Chemistry
  - Astronomy
  - Space Science
  - Microbiology and virus
  - Material science
  - Environment

- 2 Reference databases
  - China Species
  - Compound

- 4 application-Oriented databases
  - High Energy (ITER)
  - Western Environment Research
  - Ecology research
  - Qinghai Lake
Scientific Databases (cont.)

• 37 research databases
From Database to Data Cloud

• Advanced scientific data infrastructure named CAS data cloud
  – integrate all kinds of resources including various data resources, and provide cloud services for massive scientific data
  – support diversified applications and terminals to access the resources in cloud service form
Architecture and Services

Scientific data integration system **Data-HUB**

- **open data #1**
- **open data #n**

**Data-HUB** provides data services and data exchange and sharing functionalities.

**DBSpace**

- Database development platform
- Database applications

**VisualDB 3.0**

- Distributed institutional repository

- Cloud storage

**Users**

- Scientist/data manager

**Applications**

- e-Science applications
- Data integration applications

- Data-HUB enables integration and sharing of scientific data.
What we learned

• Open research data shouldn’t take scientists much time, tools and platform must be easy to use

• Open research data shouldn’t be one side and static, data sharing social network will be needed

• If data can be accessed and integrated easily, data users will become more and more, data contributors also become more and more
Conclusion

• Project-driven data sharing model is popular in China
• A clear research data sharing policy in national level and institutional level is necessary
• Tools, infrastructure, data exchange and publication model is helpful for research data sharing
Thank you for your attention!