## Open Access to the Scientific Data: Promoting Science and Technology Innovation

## Professor Guanhua XU

## Minister for Science & Technology of China Member of CAS, TWAS, IAA

20th CODATA International Conference Beijing, October 23, 2006



- I. Scientific Data Sharing and the Innovation on Science and Technology
- II. Overall Idea on Promoting the Scientific Data Sharing in China
- **III.** Progress
- **IV. Impacts**
- **V. Further Considerations**

# I. Scientific Data Sharing and the Innovation on Science and Technology

- As the economic globalization speeds up, the global flow and allocation of essential factors of production such as capital, information, technology, and talents become much more common than ever before.
- Technology advancement and innovation are turning to be the main ways of improving a country's comprehensive national strength and core competitiveness.
- It already becomes a universal strategic choice to rely on science and technology to realize the sustainable exploitation of resources and promote the harmonious development between human and nature.

# Chinese government gives a top priority to S&T development

- Compiled Development and Layout Outline of National Science and Technology Development in a Middle and Long Term (2006-2020) based on two years of strategic research
- State Department established a leadership team for development and layout of national science and technology development in a middle and long term
- Organized 2000+ experts from various disciplines and industries as participants
- Conducted strategic research on 20 key special topics and 168 key problems



## II. Overall Idea on Promoting the Scientific Data Sharing in China

## **Overall Objective (-2020)**

- Build a more user friendly scientific data management and sharing system
- Develop a set of supportive policies, laws and standards
- Build up a professional service group with career reward mechanism
- 80% of scientific data available to the general public

# Short-term Goal (-2010)

**Three-tier structure** 

- 40 scientific data centers or networks
- 300 master databases
- 1 gateway

#### Guideline

- based on the need of science and technology innovation
- integrate and open access to scientific data resources generated and accumulated by national research projects with a focus on public-welfare and basic data resources
- creating an open atmosphere for sharing scientific data resources in the society

## **Principle**

- Overall planning and resources sharing
- Cooperative constructing and unified standards
- Demand-oriented and guaranteeing security
- Priorities stand out and pilots go first

#### **Framework and Architecture**

#### **1. Logical Framework of China-SDSP**

China-SDSP is a three-tiled system: master databases, scientific data centers or networks, and Gateway Web site

#### 2. Scope of Data Sharing Supported by China-SDSP

China-SDSP also functions as a catalyst. Its original purpose is to integrate publicly funded data resources, but its long-term goal is to leverage all possible data resources from government to the private sectors, and make them available to the general public.

#### 3. Service Architecture of China-SDSP

China-SDSP may provide services in various ways: facilitating the consistent management of distributed databases; providing a content service and data service, as well as other services mentioned.



Architecture and Framework of China SDSP

## **III.** Progress

#### Policy and legal framework

- Define the legal framework of policies and regulations
  - National level
    - Science and Technology Advancement Law (revised draft),
    - Science and Technology Resource Sharing Law (research carried out on the platform),
    - Scientific Data Sharing Regulation (draft)
  - MOST
    - Administrative Method of Scientific Data Sharing Program (draft for inviting opinions)
    - Administrative Method on Data Integration/Submission of National Science and Technology Plan Projects (prepared for test)
    - Method on Selection, Examination and Evaluation Pilot Projects of Scientific Data Sharing Program (prepared for publication)
  - Participant departments
    - 39 rules and regulations on data sharing made by relevant departments in various domains

#### **Standardization**

- Develop 32 standards and criteria,23 have been completed, standard training was held for 3 times
- Based on the top design, relevant departments have made more than 120 data management standards in various industries and domains

## **Pilot projects**

- Meteorology Data Center
- Surveying & Mapping Data Center
- Hydrological Data Center
- Seismology Data Center
- Oceanic Data Center
- Land & Resource Data Center
- Agriculture Data Center
- Forestry Data Center
- Medicine & Health Data Sharing Network
- Earth System Science Data Sharing Network
- Sustainable Development Information Network

## **International cooperation**



• Study tours to the U.S. and EU countries were organized in 2004 and 2005 for exchanging experiences and making investigations on global scientific data resource management

 Scientific data centers (networks) also build channels for international cooperation and bring in large amount of global data resources

- Numbers of centers (network) cooperate with CODATA under ICSU on data exchange and sharing
- Earth System Science Data Sharing Network signed a data mirror agreement with the University of Maryland, gradually introducing satellite image data of the U.S. Landsat TM/ETM+
- With funding support from the project, meteorological data sharing center grows high potentiality for international cooperation

## **IV. Impacts**

#### Data sharing culture and data service awareness

- the concept of "Scientific Data Sharing" has been broadened from limited scientists' groups, spreading gradually among governments, departments, units and social public.
- Implementation of "the project" not only reinforced data sharing of atmosphere data within certain industry, but also drove data exchange among different industries.

# Searching "google" with the words "科学 数据共享" (Chinese characters for scientific data sharing) found over 5 Million hits.



#### Data Standards and Criteria Play Active Roles in Data Sharing in Diverse Industries

- 23 specific criteria has been completely developed, which involved in construction of scientific data resource, setting up and managing of projects, effect evaluations of scientific data sharing.
- Roles in the constructing period
  - Specified requirements for constructing and managing the project
  - Specified requirements for building up science data resources
  - Regulated the discover and access service of scientific data sharing resources

#### The Integration Revitalized Basic and Publicwelfare Scientific Data Resources

- Through SDSP, the data integration and sharing revitalized more than 25 billion Yuan of data resources produced by the state.
- According to incomplete statistics, by December 2005, pilot projects supported by SDSP had integrated and rebuilt 864 databases, more than 10 thousand data sheets and nearly 50TB data are available, which involved in almost one third of current public-welfare and basic scientific data in China.

#### **Impacts of Data Services**

- In accordance with incomplete statistics, by the end of 2005:
  - The number of accessing users: over 14 million totally
  - Registered users: nearly 50 thousand
  - Downloaded data volume:15TB

#### • User Analysis

- Over 50% of the users are from educational system and scientific research institutes
- The data obtained are mainly used in scientific research, education, science popularization and diverse plans making
- Major Projects Supported
  - Provided basic data support successively for over 1225 national key projects (of 973, 863 projects, and nature funds)
  - Effectively promoted science and technology innovation and social development in China

#### • WDC Comments

- In July, 2005, World Data Center (WDC) investigated
  9 sub centers in China. They concluded:
  - In recent years, the 9 sub centers of WDC built in China made great progress, all of which are included in SDSP
  - The amount of data continuously increased, the soft/hardware environment improved constantly, data exchanges got more and more frequent, precious materials had been successfully protected

## **Professional Teams**

Through 5 years' training, relevant departments and agencies have brought up teams of engineers. They are specialized in data standard, data sharing policy, database rebuilding and standardization, data service, data analysis, data websites. In addition, a great number of graduate students are under training.

## **V. Further Considerations**

#### Challenges

- Relevant policies, laws and regulations need to be perfected and put into practice
  - Policies to encourage data sharing, long-term effective operation mechanism and evaluation policies for data sharing, and data sharing policies of specific industries are insufficient or need to be perfected.
- The gap between SDSP general standards and specific standards in specific domains need to be filled.
  - There are gaps between SDSP general standards and specific standards in relevant domains. Some domains lack systematic data sharing standards and criteria.
- Authentic data resources need to be further integrated
  - Effective integration of dispersed data resources is necessary for building world-class databases.

## **Further Considerations**

To build up the future of China's scientific data sharing scheme, top priority must be given to 4 aspects:

- Improve laws, policies and regulations
- Promote SDSP standards and criteria system
- Construct National Scientific Data Sharing Center
- Enhance International Cooperation

## Improve laws, policies and regulations

Establish a complete policy adjustment and legal guarantee system to regulate all links of scientific data's collection, integration, sharing and utilization, adjust the interests between data owners, management staff and users, and accordingly create and maintain an order of scientific data opening and sharing, safeguard a sustainable and sound development of data sharing.

#### **Promote SDSP standards and criteria system**

Since scientific data originate from the scientific activities of observation, monitoring, research, experiment and investigative analysis from departments and units, they involve different types, complex in content, different in format and vast in amount. To realize a full sharing of scientific data, a uniform standard and criteria system must be established to guide the data conforming, integration and exchange service and improve the quality of data resources.

#### **Construct National Scientific Data Sharing Center**

MOST plans to establish a number of national scientific data centers (networks), and the national databases will be generated from these centers. The breakthrough of data resource construction shall focus on the integrated data from government produced, owned and funded projects, so as to facilitate the effective and wide use of the data by the whole society.

## **International Cooperation**

China pursues the policy of reform and opening up. We wish to introduce the mature mechanism as well as standards of data management from abroad. We encourage Chinese scientists to collaborate with international scientific community, toward a grand international platform.

#### **Future cooperation**

- Carry out exchanges, research and discussions on policies, with an emphasis on the evaluation of achievements, incentives and effective long-term operation mechanism.
- Encourage bilateral and multilateral cooperation, with a focus on international exchange and sharing of data in various domains.
- Promote training of talented people and technology exchange on policy research, standards, database technology, data analysis and network service.

## Acknowledgement

SDSP steering committee

SDSP expert committee

SDSP working group

**CODATA-China** 

CODATA

