Current Situation and Development of Data Sharing in Agricultural Science in China

Li Zhang

Agricultural Information Institute of the Chinese Academy of Agricultural Sciences, China, zhangl@mail.caas.net.cn

Abstract

With the global development of economy, science and technology, the competition of science and technology is becoming more and more intense across worldwide. In order to promote competition power, there need improved data support system. However, being social resource with broad influence and vast application value, scientific data can provide strong support to the innovation of science and technology by bring its value into full play, only if it is open to share in whole society. The paper firstly analyze the recent development of agricultural scientific data sharing in China, advance three stages of start, construction of sharing platform and application sharing service. Secondly, the paper analyze the shortage of agricultural scientific data sharing in China, put forward the following shortcomings in current sharing practice: scattered data accumulation, multi-sector management, unreasonable institution, lack of sharing mechanism, intersection of subject and complexity of sharing resources construction. However, the trend of scientific data sharing appears obvious feature of integration management, integrated application of IT and global open access. In the end, the study indicates the practical way to establish improved sharing service system, which mainly include the establishment of coordination mechanism among sectors, regulation and standard system, ensuring mechanism for long term, as well as combination with the construction of innovation system for agricultural science and technology at aspects of data resource construction, service and application. Thus scientific data bring innovation support into full play. The tasks in the future focus on drafting strategy, strengthening the publicity and raising people’s awareness of sharing, enhancing the research of technology and management, and setting up sharing steering committee etc.

Keywords: Agricultural scientific data, Data sharing, Sharing service

Introduction

Entering information era, data sharing become a prominent issue. Due to resource attribute of scientific data and its activity in increasing intense international competition, many developed countries and international organizations for science pay more attention to collection, management and sharing service of scientific data. Only if scientific data is shared among various sectors and domains, the benefits maximization of data resources can be reached. In recent years, Chinese government adequately understands the significance of scientific data sharing, and start to strongly push forward scientific data sharing. So there is much progress in data sharing for agricultural science in China. This paper will discuss current status of data sharing and its development.
1. Recent Development of Data Sharing for Agricultural Science in China

As earlier as the beginning of 1990s, Chinese scientists appealed for reform of policy to realize scientific data sharing in all society. In past 10 years, Chinese science and education community has taken active efforts to push forward scientific data sharing. In 2002, the 196th Xiangshan science conference with theme of scientific data sharing was held, which indicated that China scientific data sharing enter into new stage. In general, the development of data sharing for agricultural science in China passes through three phases.

1.1 Starting Stage (1999-2002)

Data sharing for agricultural science in China started with new science strategy of state. In 1999, Chinese government recognized the requirement of strengthening data sharing, and put forward the goal of new science strategy, i.e. strengthening infrastructure construction and open access to digital data resources. Therefore, Scientech Documentation and Information Center of the Chinese Academy of Agricultural Sciences (namely current Agricultural Information Institute) undertook the basic project on construction and sharing service of basic database for agricultural science and technology during 1999-2002 under the funding and support of Ministry of Science and Technology, P. R. C., to start strengthening the construction of basic database for agricultural science and technology. Hereafter, a lot of basic prophase works have been conducted, transferring the scattered agricultural data resources into shared information resources. Through the support of state’s project, basic database group for agricultural science and technology has been established, which constituted of 35 databases at three categories, with 1 million records and 26 GB amount of data. These data mainly include data for agricultural science and technology management, agricultural research, agricultural extension and industrialization. At the same time, selected professional institutes carried out the collection, processing work of crop genetic resource, animal nutrition and feeds, crop irrigation experiment data and corresponding database establishment.

1.2 Sharing Platform Construction Stage (2003-2005)

With pushing forward State R&D Infrastructure Platform construction, during 2003-2004 the project on construction of data sharing platform for agricultural science started to be implemented, strengthening nonprofit sharing service of agricultural scientific data. The construction of data sharing platform for agricultural science centered on digitalized integration of data and sharing service, placing emphasis on construction of technology safeguard system, data resources, data standards and sharing service system. Through the construction of sharing platform, 15 main databases and 490 databases (data sets) were integrated with amount of 54 GB data, 1 central network and 6 sub center network were established. All data were shared via internet. Under the support of sharing software system based on metadata, the sharing web station (http://www.agridata.cn) not only provide gate function for agricultural scientific data sharing, but also navigation and cross search function by integration of various kinds of agricultural scientific data using catalogue database of agricultural metadata.
1.3 Discussing Sharing Application Service Stage (2006-)

Based on the previous two stages, data sharing for agricultural science in China was stably advanced. In recent years, the sharing application service was actively discussed under strong safeguard of data resources, technology, facilities, staff, finance and institution. Through study and practice, the following three modes were applied to provide sharing service. The first is for digitalized data, providing sharing service through metadata; the second is for non digitalized data while without conditions for database establishment, providing sharing service through metadata by utilization of spatial remote technology for database establishment provided by national agricultural data center; the third is for non digitalized data and those data can not be networking sharing in near future, applying the method of registering service information to make users find data owners by query of service information and realizing data sharing by real means. At present, the application mechanism of sharing service for agricultural data are further studied.

2. Defects Analysis of Data Sharing for Agricultural Science in China

2.1 Scattered Data Accumulation and Multi Management Sectors

The current status of production and distribution of agricultural scientific data in China is as follows: a) collection, management and maintenance of scientific data are completed by government’s investment; b) data resources scatter in various sectors, institutions and individuals; c)data accumulation show cutting apart pattern of sector within and across industries. Data resources mainly distribute in three systems of public research institutions, higher agricultural colleges and agricultural enterprises. The overall situation of data resource distribution complies with the distribution rule of agricultural research institutions, colleges and universities. Agricultural research institutions, colleges and universities at central level possess more types of data resource with quantity of comparable magnitude and better quality than those at local level. Except for the selected data resources on agricultural basic discipline owned by Chinese Academy of Science, agricultural scientific data resources owned by the Ministry of Agriculture is richest, with high quality. At the same time, the scattered agricultural data belong to various sectors under current management system. So the extent of centralized management for data resource is relative low. Data management is conducted by various sectors in their own ways. Due to this, there are still more difficulties in reaching data sharing.

2.2 Unreasonable Institution System and Lack of Sharing Mechanism

Under the influence of current system, scientific data for agricultural science is subordinate to various sectors. Even for data in special domain, it is managed by different sector and institution respectively, and it is preserved and controlled by different research group and researcher within sector and institution. In view of individual’s benefit, each management department or individual develop correlative stipulations on preservation and management of data resources, which are not unified, even contradictive. Therefore, there lack communication and coordination mechanism among sectors, so as to cause repeat and wasting of data resources. The lack of awareness for cooperation among sectors brings difficulties in forming social join forces, which doesn’t help to exert macro-control effect of
state to improve investment efficiency. In other words, such unreasonable system leads to lack of main body for data sharing service. Data sharing is not subject to macro-coordination and management of state. On the other hand, due to lack of feasible sharing mechanism, including data management, data collection and service mechanism, the data from research project funded by government and public research institution is not timely, effectively developed and utilized.

2.3 Discipline Intersection and Complexity of Sharing Resources Construction

The development of agricultural science trends to comprehensive cross subject, therefore the collection and accumulation of agricultural data exist overlapping. The overlapping resources are unavoidable in resource construction and sharing. For example, some of agrometeorological data, agrological data, agrohydrological data, agromicrobiological data, agroecological data etc. are classified as meteorological science data, geological science data, hydrological and water resources data, as well as data for science database of Chinese Academy of Science, with realization of sharing via correlative web station. In addition, agricultural germplasm resource data (including crop, animal and microbe) is integrated to information network for Chinese sustainable development and biological information network, monitor data for agricultural environment to environment protection network, agricultural biotechnological data to life science network. For this reason, the construction and sharing of agricultural data is related to sharing of scientific data in relevant discipline. Joint cooperation and data sharing in multiply ways is necessary. Besides, the data resource construction is complex and involved in united integration of many techniques in different domain at various levels. This present vast challenge on large-scaled integration of data resources and standardized sharing.

3. Characteristics of Development Trend of Data Sharing Management

3.1 Integrated Management

Scientific data sharing is a reform of data resource management. Realizing sharing imply digitalization of data resources, digitalization, networking and intelligence of data management, communication and transmitting method. Under traditional technology condition, sharing of data resource indicates the sharing of data carrier, i.e. sharing of documentation. In the traditional sharing system, the provider, producer, distributor and user are independent each other. It is similar to various sector of industrial chain, responsible for data creation, processing and utilization respectively. Under internet environment, the role of all participants is not clearly divided in new type of data sharing management pattern. Data provider maybe data processor and user, so the data quality of sharing system is effectively ensured. The benefit of all participants in such interactive sharing is balanced. Ethics, good faith and expertise provide guarantee for data quality. Therefore, data sharing is characterized by integrated management during process of data creating, processing and using.

3.2 Integrated Application of IT

The construction of data sharing system for agricultural science not only require enhancement of the construction and integration of digital data resource, but also guarantee
of strong technology and network support system, with driving force of the innovation of system and mechanism. As dominant technology of sharing, ITs entirely change the features of data and its circle pattern. The detailed influence is as follows: a) database technology brings micro and electronic data carrier, realizing automation of data management; b) multimedia technology brings the integration and interaction of multiply data; c) information highway and modern communication technology make high efficiency, speed and networking of data transmitting. Therefore, the establishment of data sharing platform by integrated application of IT provide technological support for function realization of sharing system. Data sharing platform provide comprehensive and deep resource sharing by integrating mass data in distributed and heterogeneous database using IT, which include network, data and service platform. In addition, integrated service of sharing requires technological foundation for environment of digitalization, networking, intelligence and virtualization. For the above reason, integrated sharing technology is also main support to sharing, including technology for metadata database, heterogeneous database, networking database etc.

3.3 Global Open Access to Data

Since 1950s, the development of data sharing tends to international and global trend in order to adapt to demand from global research. In view of international development for data sharing, there are prominent features in accumulation, storage and utilization of scientific data. It is basic task of state; the data from all kinds of observation station funded by government is shared; the implementation of data accumulation and research plan is synchronous; the data from important research project funded by government is provided for sharing after project executive analyze and study for two years; unified standard is developed and theory and method study is carried out; the high technology is introduced to facilitate utilization of data, with sufficient role in promoting science advance, social and economic development. In general, data service at international level is provided according to particular regulation and unified standard in way of non-commercial, non-profit, free exchange, grading and classification. International experience in data sharing presents salutary lesson for China.

4. Practical Approaches to Establishing and Improving Sharing Service System

4.1 To Build Mechanism for Coordination among Sectors

At present, most of data resources are respectively preserved in professional institute of various sectors with decentralized management. But the same unit maybe play role of data preservation and data utilization at same time. Although data resource is asset of state, sharing behavior is driven by benefit during practical operation, which does not contribute to equitablity and efficiency of data sharing. Therefore, establishing management organization for agricultural scientific data sharing may enhance coordination to assure the creation of smooth path for shared information communication. By building mechanism for coordination among sectors, the moderate separation between data owner and manager can be made with the guarantee of equitable and justified sharing in manner of third party. In addition, it helps realizing overall plan, coordinated development to avoid paying more attention to data collection and storage than data sharing service by various sectors. It will
change the situation of lack of organization responsible for sharing service and relevant promise.

4.2 To Develop Regulation and Standard System

Scientific data sharing is involved in aspects of data property right system, encouragement and supervision mechanism, technology decision, safe and performance evaluation mechanism, so agricultural scientific data should belong to state property according to the above legal framework that need to be established. The responsibility, right and benefit of data registered unit should be identified to push forward data resource sharing within or cross sectors. The measures include the development of administrative regulations and local rules. At the same time, policy making for submitting of data from research project is also necessary for advocating integrated management of data resource. On the other hand, in order to promote utilization level of digital data and service capacity for integrated sharing, a set of standards should be developed during the digitalized construction and integration of agricultural data. The government may attract department with work basis and capability for research and development to take part in the study and formulation of relevant technological standard and data standard, by bidding and entrusting. And the standards need to be increasing established and improved. Government should play dominant role in dynamics maintenance, expansion, publicity and extension of the standards.

4.3 To Build Long-run Safeguard Mechanism

Data sharing is a complex systemic project, requiring providing guarantee in data resource construction, finance, technology, staff and facilities. Therefore, the establishment of long-run safeguard mechanism should consider the above aspects. For resource construction, coconstruction and sharing is a feasible way. Nowadays various departments conduct collection and digitalized processing of data, which is decided by feature of agricultural science research. So, the coconstruction mode by all society with main investment of government should be applied. The government guide data processing of data origin unit by encouraging policy, with activity of sectoral and local participation. Scientech resource is refined disposed by unified plan to strengthen integration and coconstruction. In finance aspect, the fund of sectoral and local construction can be gained in different ways which include state finance, local finance, investment of sector and enterprise. As regarding the funding of sector and enterprise, government should develop preferential policy for compensation. At the same time, government should expand special investment on digitalization and integration of data resource. In respect of technology, staff and facilities, research and development of technology, talent team construction, infrastructure construction including equipment should be enhanced. In addition, the supervision and evaluation mechanism should be established to ensure the sustainable development of sharing.

4.4 To Combine with Construction of Innovation System for Agricultural Science and Technology

The current situation of data sharing for agricultural science is close to system of agricultural research, so resource construction, sharing service and data application should comply with the establishment of innovation system for agricultural science and technology.
According to the three-level network system of science and technology innovation, data innovation system should be established with national data center, regional data center and data center for experiment station. The dynamic and open data system can be ensured by lateral, vertical data correlation and flow. The new type of management mode for data sharing reach scientific data and knowledge sharing by data service of internet, resource portal, visual library, to promote circle, recommendation and distribution of scientific data, as well as distribution and education of knowledge, transformation and extension of innovative achievements.

5. Emphasis of Data Sharing Work in Near Future

5.1 To Develop Strategy of Sharing Development

The formulation of sharing development strategy can provide effective guarantee for advancing the high-efficient disposition and integrated utilization of scientific data resources in all society and enhancing support capacity for innovation of science and technology. In the future, the digitalization and integrated sharing of agricultural scientific data should be incorporated into national medium and long term development plan for agricultural science and technology to enhance the macro design and management. By strategic plan to push forward the construction of all kinds of sharing service network oriented various innovation entities, the national data sharing service system is established with grading and classification of users. At the same time, the capability of exchanging information with international data organization should be promoted by reasonable plan of constructing scientific, improved data system for agricultural science.

5.2 To Strengthen Propanda to Popularize Sharing Awareness

Artificial obstacles in scientific data sharing mainly depend on the awareness and willingness of researchers. Since a long time ago, the personal awareness of monopolizing scientific data is relatively strong and there lacks cooperation sprit. Therefore, the sharing understanding of more researchers should be improved by policy guide, publicity and education. By actively advocating the attribute of state intellectual property for scientific data, the benefit of state is strengthened for ensuring that benefit of department and individual comply with benefit of state. The rule of intelligent property right protection should be appropriately applied by breaking through the boundary of intelligence property. According to various categories of scientific data, different sharing strategies should be applied to eliminate the researcher’s worries. By the demonstration of data sharing network, manager firstly challenge tradition and change management pattern with establishment of encouraging mechanism. In formulation of policies on shared data content, service methods, sharing organization development and fund support, the participant method should be adopted to sufficiently arouse the activity of all participants in social sharing for forming of sharing atmosphere.

5.3 To Enhance Study on Technology and Management

With the development of science and technology, the demand for scientific data is increasing. It is necessary to gradually advance the progress of sharing technology and reduce operation cost of data and system, so as to ensure more users apply sharing system and data, while
promote data product and service capacity of agricultural scientific data. There need to
strengthen the infrastructure construction of data sharing to sufficiently apply IT to promote
data sharing. The innovation of sharing technology should be further push forward,
especially technology for data discovery and capture. For example, implicit knowledge in
data can be mined and acquired by application of data warehouse, document management
system, management information system(MIS), data mining, expert system and artificial
intelligence technology; sufficient knowledge sharing can be reached by application of
network technology, groupware technology, knowledge map, CSCW, new type retrieval
techniques and so on. While enhancement of technology research, new mode for sharing
management should be further studied to continuously reconcile all kinds of contradiction
with which sharing management faced, and timely resolve new problems in sharing
management.

5.4 To Establish Steering Committee for Data Sharing

In order to adapt to development of information era, it is necessary to enhance digitalized
processing of data resources in hard copies, centralized development and utilization. For this,
there need to establish work operation system for coordination management of data
resources construction and integration with government dominant and multi-sector
participation. By the establishment of steering committee for data sharing to strengthen
organization and management, the work operation system is improved to provide strong
organization guarantee for realization of data sharing. The introduction and coordination
group constituted of leaders from government and related sectors need to be established at
central government and local level, which is responsible for coordination work on
digitalization and integration sharing across multi-sectors, and policy development. The
group will advance overall plan and distribution recommendations of data resource
digitalization and integration, to push forward related work on construction and sharing of
agricultural scientific data resources.

References

China Agricultural Science and Technology Press.