



The Countermeasures of Carrying on Web of the Research Institutions in the Era of Big Data - Consider the Web of Chinese Academy of Agricultural Sciences

Liping Zhang

► To cite this version:

Liping Zhang. The Countermeasures of Carrying on Web of the Research Institutions in the Era of Big Data - Consider the Web of Chinese Academy of Agricultural Sciences. 9th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Sep 2015, Beijing, China. pp.382-391, 10.1007/978-3-319-48354-2_38 . hal-01614213

HAL Id: hal-01614213

<https://inria.hal.science/hal-01614213>

Submitted on 10 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

The Countermeasures of carrying on Web of the Research Institutions in the era of big data

---Consider the web of Chinese Academy of Agricultural Sciences

Liping Zhang*

Institute of Agricultural Information, Chinese Academy of Agricultural Sciences,
Beijing 100081, China Key Laboratory of Agricultural Information Service Technology
(2006-2010) , Ministry of Agriculture, The People' s Republic of China
zhangliping@caas.cn

Abstract. In recent years, large data caused great concern in industry, academia and government. As an important department for scientific research and innovation, the academy portal show the level of science research innovation ability and it is a main platform of transformation of scientific research achievement. In this paper, the construction portal of Chinese Academy of Agricultural Sciences Situation as the background, points out the shortcomings of the current construction site, and make a few suggestions for website development in the big data environment.

Keywords: Big data; Website construction; data resources

1 Background

Big data is hot word in today's society of network . People from ordinary to the whole country enjoy the convenience of big data in varying degrees. In medical services, retail, finance, manufacturing, logistics, telecommunications and other industries, the research and application of big data has started, and has created a huge social value. Government departments also attach great importance to large data technology. In March 2012, Obama announced the U.S. government to invest 2 billion dollar for the big data research and development program. Our government has also realized the huge potential of big data and is to make policy to promote the development and application of big data.

With application of information acquisition, mass data storage, high-speed data transmission and intelligent data analysis, it has brought a far-reaching impact on the contemporary agricultural scientific research. Chinese Academy of Agricultural Sciences (hereinafter referred to as CAAS),as the most authoritative scientific research institution in the country, have already felt the great influence on big data. In 2014 , CAAS draw up "China Academy of Agricultural Sciences information development planning (2015-2019)" (hereinafter referred to as the planning).The planning put forward that the main task of is to build e-Science platform, promote

¹information of scientific research conditions, management and scientific research output and enhance capacity for the scientific data and resources acquisition, protection and utilization. The acquisition and utilization of scientific research data will be an important work in the future. At the same time, the planning points out the problems in the construction of data.

2 current situation and problems

CAAS Web Portal included 10 functional departments websites and 41 research institutes websites ,have two major functions of department office and service research, and it is an importance display window for showing scientific research strength and service agriculture. Since it was on the line in 1997, website has played an important role for science research and management. Now, CAAS portal has included 51 institutes secondary network. But according to statistics, total websites in CAAS has arrived 236. The specific site condition survey results are shown in table 1.

Table 1. Number of existing sites and views

Unit	The number of sites	IP average daily traffic	Average daily page views
Agricultural Information Institute of CAAS	43	1310	7647
China National Rice Research Institute	19	1009	4211
Food Science and Technology Institute of CAAS	13	1010	3123
Farmland Irrigation Research Institute	13	977	2785
Management of CAAS	12	18901	98142
Biotechnology Research Institute of CAAS	12	812	4205
Institute of Environment and Sustainable Development in Agriculture,CAAS	11	657	1223
Biogas Institute of Ministry of Agriculture	11	612	1129
.....

Construction of Web database has begun in the 1980' s and it has arrived 185 as of 2014.Here are several examples in early phase. In the 1980s, the database of feed nutrition had been jointly build by Institute of Animal Sciences and the Computing Center head by Academician Ziyi Zhang.In the 1990s,seed resources database built by professor Xian Zhen Zhang has reached 42 million types by 2014. The agricultural

¹ Foundation item: Chinese Academy of Agricultural Sciences website content resources, research organizations and service models

science and technology data platform information presided over of a few years ago by professor Xianxue Meng had been put into application in which involves 12 individual sciences, 72 subject database groups, 672 databases and 1987TB data quantity browsed by 50-60 million people every day. Seed resources database in China has accumulated hundreds of millions of seed resource data. With development of genetics and breeding subject the gene sequence of important crops and animals measured precisely have formed a huge gene database. There are many All kinds of scale national construction scientific data and resource data center. But, it has a weak ability in data collection and integration. There has only 14 data websites in the CAAS portal. According to statistics in 2014, the other 185 data resources websites are independent with the CAAS portal. All of these data resources can't share and analysis. Web site data storage format and location are shown in table 2.

table 2: Web site data storage format and location

Number of sites	Data storage form		storage location		
	database	spreadsheet	network center in CAAS	network in institutes	Network in The third party
236	227	9	21	22	193

2.1 Network is weak

2.1.1 Lack of overall website construction planning and financial support.

Since the CAAS portal has been setup, website establishment, development and daily maintenance run by each institutes. Because there is no unified planning and special funds to support, website construction and service level is uneven and database construction and maintenance status are poor. Nothing of unified management and data sharing platform it is difficult to data sharing and analysis in the future.

According to statistics, nearly 2/5 of the institutes do not have the operation and maintenance expenses. Compared with the Chinese academy of sciences, CAAS funds in development of information is far smaller. CAS information construction began in the "fifth" period. By the end of "Thirteen Five", it has accumulated invested several million dollars.

2.1.2 Network Infrastructure environment is weak.

According to statistics, Chinese Academy of Agricultural Sciences 41 research institutes, 10 departments in 2014 generated approximately 10TB of data per day, and 10% annual growth rate, while the storage capacity of the existing network center room of the CAAS can not meet demand. All of the institutes in CAAS distributed 24 provinces in the country. Since there is no interconnection of data transmission lines,

scientific research data between the institute, the base, the test station and field station can not be shared and transmitted.

2.2 Lack of unified planning of data resource construction

Data resource construction is a long-term work and need a reasonable system to ensure data accuracy and availability. lacking of unified standards data need to repeat collection, repeat input and maintained by multiple departments. At it caused information update asynchronously and non unified. The specific problems in following several aspects:

First, there is not a unified data format and compatible standards. According to the survey, all of 42 institutes have built their own financial and personnel platform, but because CAAS did not establish a unified system compatibility standards, data in one departments is not compatible with another department. As the amount of data is growing rapidly, it will increase the workload of management and work harder.

Secondly, CAAS lack research management platform, such as platform for research software and scientific instruments sharing. Instrument and software is an important part of the use of scientific research funds. Because CAAS have not scientific research software and instrument unified purchase plan and sharing mechanism it caused software and instrume purchased repeatly and utilized lowly.

Third, there is a lack of scientific research data storage mechanism. Most of the scientific research data edit and save by each group, lacking of system and professional. According to statistics in 2014, 115 research institutes or units has built its own web site that one institute have about 3.3 of web sites on the average. But it is lack of unified management and effective utilization , scientific research data is low usage rate .

2.3 Lowly data processing capacity

With the requirements of computing resources and storage resources continue to improve, high performance data processing platform is necessary condition to deal with data resources. The development and use of high performance computing in 80 years of the last century has already begun. National "973" and "863" programs had put on a lot of funds in research of high performance computing and it had applied many fields such as defense and security, oil exploration, weather forecasting, bioinformatics, gene and nanotechnology and other aspects. In "fifteen" period, CAS has made remarkable achievements in high-performance computing environment and had developed domestic-made supercomputer Shuguang 2000-II and Lenovo Pentium 6800 supercomputer. But CAAS has not its own high-performance computing center and all high-performance computing demands are dependent on other societies. Since the high-performance computing research in different directions have different software and hardware requirements and highly specialized, the results computed by outside computing center cannot reach the preset requirements, and data security can't be protected. The survey in 2014 showed that more than 2/3 institutes particularly animal husbandry, biology, vegetables and

flowers, have used high performance data processing, while other institutes also expressed requirement of high performance data processing.

2.4 website services model is single

In the current era of rapid development of the Internet, the ways people access information and communication are also diverse, new media such as Twitter and micro-channel has become an important way of communicating information. And smart phones, tablet PCs and other mobile Internet terminals grow explosively. As of mid-June 2013, China's mobile phone users reached 464 million. Mobile office has become a more relaxed and effective way of working in the future. But CAAS portal website only has a traditional Internet platform and is lack of function to diverse and release data resources.

3 Suggestions of web data resources construction in the future

The planning in 2014 proposed that information technology development goals in the future will be around the overall goal of "build a world-class agricultural research institutes," and build domestic first-class agricultural research institutions in information service platform and to provide first-class information technology services for agricultural research through the integration and sharing of agricultural research data.

3.1 Improve system of data resources construction

Construction of the database is a long-term project. Ensuring the smooth development of construction of the database, we need to establish clear rules and regulations in the fund, personnel and data collection, storage and so on.

3.1.1 Have a realistic implementation plan and funding.

The planning in 2014 has listed a timetable implementation for specific information construction which including planning storage and computing platforms and data resources construction funds needed. Program is divided into two phases, the first five years will focus on the construction of network infrastructure, and the second year will deepen the service and data mining. The first task of a five-year plan have been identified, the specific implementation plan table 3.

3.1.2 Construction of the data resources needs full-time institutions and technical personnel.

We should establish academy-level information technology allied agencies and improve the information technology systems of each institute. Currently, information construction in CAAS is responsible of Agricultural Information Institute. Only two

related offices including 25 employees are charge of network maintenance and site updates. Other information technology personnel is even more lacking. In the planning, CAAS will established a-hundred-person team of network operation and maintenance and website editors ,and set policies to improve staffing and operational system of information technology ,and regular train and assessment technical personnel. Hence we need to establish comprehensive information centralized management and to format a clear division working mechanism for each institute. And on this basis we would explore to establish the academy co-ordination, each institute the participation, cooperation and win-win management and service models.

Table 3:The planning timetable from 2015 to 2020

No	program	aim	Execution Time
1	Rules and regulations	Formulate technology standards framework	2015.1-2015.6
		Formulate Information construction standards	2016.7-2017.12
		Formulate operation and management of information	2015.7-2016.12
2	Administrative Information Construction	Institutes information database based on completed and e-government applications.	2015.1-2016.12
3	Network-based environmental improvement	Construction of high-speed and stable network transmission channel and network bandwidth access levels and school district Wireless network coverage	2015.1-2016.12
4	High-performance computing platform	Construction of high-performance computing platform to achieve effective mining resources, integration, sharing.	2018.1-2018.12
5	Construction site of new media platforms.	Construction of multi-media platforms and carrier information dissemination, diversify information services.	2015.1-2019.12
6	Security conditions building	Through third-party security cloud platform, to all sites within the Academy belongs range safety monitoring service.	2016.1-2019.12
7	Data Sharing and Analysis Platform	Scientific Data virtual shared environment construction	2016.1-2017.12
		Data analysis platform (tools) Construction	2017.1-2019.12

3.2 To improve network infrastructure construction Environment

Network infrastructure is a prerequisite to protect the data storage and computing preconditions. In order to meet the data resource construction of network infrastructure environment we should know the needs of each network. First, it is necessary to clearly know network requirement and to overall layout design network

function, structure and layout. Meanwhile we should improve power system, monitoring systems, fire systems, refrigeration systems and other facilities. The Planning put forward that CAAS will be completed in the network infrastructure to meet the next decade, the development of information technology within five years(Table 4);Secondly, we should construct high-speed data transmission channels, improve network transmission environment and establish a without barriers data sharing and transmission line to link 42 institutes distributed in the 20 provinces.

Table 4: 2015 - - 2020 the central office-based environment construction plan

program	aim
Center room area	By 2020, the total area of 5000m ²
Power Systems	Build double circuit power supply systems and power protection UPS8 hour online
Security System	The establishment of advanced technology and personnel management of access control and fingerprint input facilities
Surveillance system	Establishment of an international advanced automatic monitoring system, electricity, air conditioning, UPS automatic monitoring
Fire Fighting System	The establishment of smoke, temperature sense of alarm systems and automatic fire extinguishing equipment
Cooling System	Using professional precision equipment to ensure the room temperature, humidity, fresh air status

3.3 To build network platform of the development for data resources

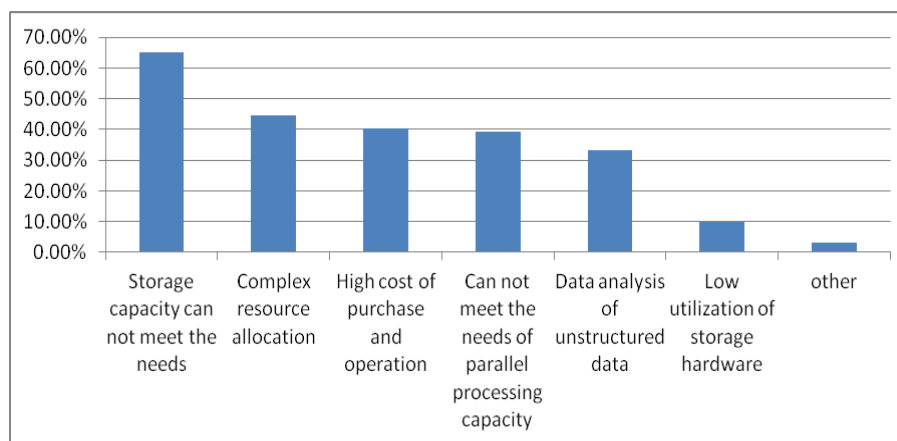
Data resources are one of the main research achievements of scientific research. According to research firm IDC predicts that the world's raw data storage capacity will be an annual increase of more than 50% by weight, and all of data is not only growing in volume but the growing complexity of data resources. Therefore, the establishment of massive data collection, storage platforms and computing center will be a necessary condition for the era of big data research institutes and development.

3.3.1 Construction big data cloud storage center

Big data analysis and research has been carried out and quickly deployed in a variety of different research areas, such as genomics, proteomics in particular, its data growth rate will exceed the legendary speed IT design development, because data storage capacity and data processing capacity of the existing data center can't meet the future needs of scientific computing and analysis. Survey about information technology showed in Figure 1: This indicates that in the case of rapid growth of data the pressure storage capacity is the biggest problem of the network construction. Therefore, in order to meet the data storage demands of big data era, building cloud storage center is to meet the necessary conditions for the future of store large data.

The planning proposed to establish a data cloud storage platform for information integration and islands of information and proposed to provide long-term data preservation service, remote backup service and online storage service associated monitoring and surveillance data.

Figure 1: Chinese Academy of Agricultural Sciences Research Network Storage Problems table



3.3.2 Construction High Performance Computing Data Center

In contemporary agricultural science research, since scientific data surge that the possibility of obtaining more and more science depends on acquiring, processing a sufficient amount of data capacity. Establishment of high-performance computing data center can provide genomics, proteomics, bioinformatics, new materials and other high-performance computing services.

CAAS have a strong demand for high-performance computing. According to 2014 survey results, more than 2/3 of the institutes is applying high performance computing, 1/3 of the institutes also has the needs of high performance computing. Along with the agricultural development of information technology, agricultural research is becoming comprehensive and interactive. It is urgent to the needs of multi-field, regional, cooperation in collaborative research teams. All types information collection and data mining and analysis has become the main direction of agricultural research and development. Because the database has accumulated a massive data of flora and fauna resources, monitoring and sensing data, network data, traditional data mining massive data mining model can't meet the demand for computing power, thus it is the need to establish a new data mining models with high-performance computing capabilities.

Building high-performance computing centers should combine with the actual needs of research work. By the research institutes requirement of the high-performance computing application, we can build a high-performance computing platform including a scientific-basis parallel computing platform software platform, system application software platform and tools.

Build high-performance computing centers can also provide remote sensing data, sensor data, network data and other large data analysis and processing business for

domestic and international users. Building cloud computing platform will achieve effective dispersion mining resources, integration, sharing.

3.3.3 Construction of the data collection and processing center resources

Website statistics show that in 2015, all of the 236 sites, 227 sites generated data used the database management. Data collection and processing center construction can rely on the network resource data platform to meet the decentralized or centralized storage of agricultural natural resources and science and data resources to achieve centralized remote integration and data sharing agricultural scientific data resources. Construction of Agricultural Sciences data integration center, we can improve the efficiency of data query services, protect the security of the data.

For experimental data (experimental) chamber of agriculture science, agricultural science field station and observation station generated, we can deploy data collection layer, field control layer, data storage layer and business application layer architecture system, establish a wireless sensor network system, the completion of agricultural production monitoring network laying demand, eventually things will transmit information to the data collection center.

Ultimately, by constructing a data center networking platform we can provide experimental (test) data network platform for intelligence gathering and storage, and improve the processing efficiency of field observation data.

3.3.4 Established national agricultural resources Things monitoring platform

Establishment of national agricultural resources monitoring platform can integrate multi agricultural condition monitoring resources, expand and upgrade the existing infrastructure integration and software integration systems. According to the principle of cloud computing and cloud service management, we can design and build the platform architecture, deployment platform equipment, development software system, then gradually establish a national agricultural condition monitoring platform. The platform can achieve coverage area of the county's main crop types, to provide real-time dynamic for agricultural research and agricultural production and long-term accumulation of data resources, to provide support for agricultural production and scientific guidance for disaster emergency management.

3.4 broaden the web services model, increased resource sharing and communication platform

The planning in 2014, has put forward to use the latest technology such as responsive web design to optimize the design and development of adaptive Chinese portals, and develop the network carriers to support for smartphones, tablet computers, TV, PC monitor, IOS and Android mobile phone carrier access.

New media platforms is another task, such as building weibo WeChat which can broaden the personalized information service mode and active push, diversified information services.

4 Conclusion

A large-scale production, sharing and application data era is just beginning, as Professor Victor said, the real value of big data is like an iceberg floating in the ocean, we can only see the tip of the iceberg at first glance, most hide beneath the surface. Carry out research work, will also be with the arrival of the era of big data, a huge change, and construction site data resources, but also will become all the researchers involved in scientific research is an important content.

References

1. Viktor Mayer-Schonberger ,Kenneth Cukier. Big Data [M]. Zhejiang: Zhejiang people's Publishing House,2013: 127-156.
2. Huang Ruhua ,Wang Bin,Zhou Zhifeng. A Study on Countermeasures for Promoting Scientific Data Sharing in China [J]. Library, 2014, (3):7-13(in Chinese).
3. Xu Zhejun, Fu Yao. Exploration of Scientific Research Management Informationization in Context of Big Data. [J] CHNOLOGY AND INNOVATION MANAGEMENT,2014,2:112-115.
4. Zou Shifang. Application of Personalized Recommendation Technology in Agricultural Extension System-Based on the United State department of agriculture website case analysis.[NH] Central China Normal University, 2014:10-14.
5. Wen Yuan.Research on Internet Text Mining and Personalized Recommendation .[NH] Beijing Jiaotong University.2014:18-24.
6. Feng Xixi, Zhu Xuefang. The Design and Implementation of Information Resources Extractor of Agricultural Websites Based on Spring[J].Information Research,2011,3:19-22.
7. An Yue,Li Bing,Yang Ruitai,Hu Lidan. Content-based Personalized Recommendation on Popular Micro-topic[J]. Journal Of Intelligence,2014(2):155-160.
8. Xia Qiuju,Li Wenchao et. New Social Networking Platform for the Academic Fields:Research Networking Systems[J]. Journal Of Intelligence,2014(9):167-172.
9. Zhao Yi,Zhu Peng,Chi Xuebin et.A Brief View on Requirements and Development of High Performance Computing Application[J]. Journal of Computer Research and Development,2007(10):1640-1646