

Research on the Cultivation of Research Data Service Talents for “Double first-Class” Construction

Pingping Yu

Huizhou University Library, Huizhou, Guangdong 516007, China

ABSTRACT. “Double-class” construction is the strategic goal and task of China's higher education development in the new era. The cultivation of innovative and applied talents is one of the missions of “double first-class” construction. Under the background of “double first-class” construction, the cultivation of talents in research universities in China should be used for reference. The successful experience of foreign top universities, clarify the research functions of research universities and modern universities, coordinate the relationship between teaching and research, realize the transfer of academic disciplines and scientific research advantages to undergraduate teaching and talent training, focus on the combination of production, study and research, and promote domestic universities. Exchanges and cooperation with top international universities. The guiding ideology of implementing innovative talent training objectives and innovative talent training system is an important issue to be solved in the current “double first-class” university construction. This paper based on the “Double first-class” construction background, analyzes the status of scientific research data services and the ability of the job, discusses the problems existing in the training of data talents in China, and proposes paths for cultivation of research data service Talents, which is conducive to giving full play to data strategic resources and decision support functions and providing powerful data support services for “first-class”.

KEYWORDS: “Double first-class” Construction; Research Data Service; Talent Training

1. Introduction

In October 2015, the State Council issued the “Overall Plan for Coordinating the Advancement of World-Class Universities and First-Class Disciplines” and proposed a major strategic decision to “build a world-class university and a first-class discipline”. The “double first-class” construction is a leap-forward development of China's higher education development strategy, which is conducive to comprehensively improving the international status and higher education quality of China's higher education. General Secretary Xi Jinping emphasized in the collective study of the Central Political Bureau. Promote the implementation of

national big data strategy, accelerate the improvement of digital infrastructure, promote the integration and open sharing of data resources, ensure data security, and accelerate the construction of digital Chinese scientific data is a new production factor, a basic resource and a strategic resource, and also an important productivity. It will change the scientific research paradigm and the knowledge exchange ecology, enhance the ability of scientific and technological innovation, and promote economic and social development. In the context of big data, scientific research has shifted to a data-intensive scientific research paradigm. "The design and implementation of scientific research activities will be around data, and the collection, filtering, calculation, storage and sharing of data will become the subject of scientific research. Data has become the basic unit of academic exchanges between scientific research institutions and researchers." (Huang Xin, 2016). Data is not only the result of scientific research, but also the object and tool of scientific research. Chinese universities are also constructing data-driven disciplinary development systems and decision support systems, more and more relying on the management and analysis of massive subject data, and urgently needing to strengthen the training and construction of research data service talents. Based on the "Double first-class" construction background, this paper analyzes the status of scientific research data services and the ability of the job, discusses the problems existing in the training of data talents in China, and proposes paths for cultivation of research data service Talents, which is conducive to giving full play to data strategic resources and decision support functions and providing powerful data support services for "first-class".

2. "Double first-class" construction, data talent training and scientific research

2.1 The relationship between "Double first-class" construction and data talent training

Data talent training is the inherent need of "Double first-class" construction. The "Double first-class" construction integrates China's higher education into the world-class higher education system for global comparison and competition, and be competitive with the characteristic advantages, prompting Chinese universities to launch a new round of intense discipline construction and talent competition. Whether it is the introduction of subject talents or the analysis of subject competitiveness, it is necessary to use data as the basis, and to use scientific research data analysis reports or disciplinary development evaluation reports as the basis. Faced with massive scientific research data, the management and analysis of university research data is imminent. At present, the groups that participate in scientific research data management in the field of higher education mainly include university libraries, information technology service departments and research support departments, and cannot fully meet the data management needs. Library data librarians have a strong background in intelligence knowledge, but often lack data technology; data technology analysts have strong data technology, but often lack the professional background of disciplines; scientific research managers have

the business ability of scientific research management, but lack of data Technical and intelligence analysis capabilities. Therefore, cultivating data talents with comprehensive application capabilities is a need for “Double first-class” construction.

Data talents training is also one of the goals of “double first class” construction. At present, with the massive growth of data and the recognition of the value of data, the importance of data management, mining and utilization has been recognized in various fields of various industries. However, effective management of data requires professional data processing capabilities, which is a major challenge has led to an increasing demand for data management professionals, which is in short supply. In 2017, McKinsey predicted the data talent gap of 1.7 million in 2018 in the Data Science Industry Salary Report. This urgently requires universities to vigorously cultivate data talents. Under the background of “Double first-class” construction, colleges and universities can enhance their international status only by strengthening the cultivation of “first-class” application-oriented data talents and continuously meeting the needs of social development.

2.2 The relationship between talent training and scientific research

The core goal of “Double first-class” construction is to cultivate first-class talents and produce first-class scientific research results. The cultivation of first-class talents is mainly aimed at students' teaching, and the first-class achievements are mainly aimed at teachers' research, which are complementary. The construction of first-class disciplines in the new era depends on the two-way support of scientific research and teaching. The foundation of the “Double first-class” construction lies in the construction of first-class courses system, and the cultivation of first-class talents is inseparable from the support of first-class scientific research. The main product of scientific research is the knowledge of innovation, which is the premise and foundation of the sustainable development of teaching, and scientific research itself is a higher level teaching process. Teaching is also an important source of power for scientific research. In the process of building a high-level university, we must deal with the relationship between “research and teaching”, and promote the coordinated development of disciplines, majors, and teaching work from concepts, systems, and initiatives.

3. Overview of Research Data Services

3.1 Basic concepts of Research data service

With the advent of the big data era, the advantages of data as a new strategic resource have become increasingly apparent and become the main driving force for innovation. In the context of data-intensive research, research data has become an important asset supporting the development of scientific research. Research Data, also known as research data or scientific data, is the original material for recording

facts collected, created, and retained by researchers to produce or validate scientific research, such as observation, experimentation, derivation or compilation, simulation, and reference. Or standardized scientific process data, semi-finished products or results data. Scientific research data is an important part of academic output and an important reference source and driving force for academic research.

Regarding the concept of research data services, there is no unified definition at home and abroad. The Australian National Data Service believes that the understanding of research data by different organizations will be affected by the environment in which they are located, so it is difficult to clearly define the research data. Some scholars believe that research data services, scientific data services, research data management services are the same concept and many forms in the expression, the core of which is to assist users to carry out scientific research and realize the existing value and added value of scientific data; Some scholars have also analyzed the concept of research data and scientific data. They believe that the connotation and extension of scientific data are broader than research data. Research data services and research data management services have different focuses. Scientific data services cover scientific research data services and scientific research data management services. Cox (2016) believes that research data management includes “a range of activities and processes in the research life cycle, including data construction and generation, storage, security, preservation, sharing and reuse, as well as technical, ethical, legal and regulatory issues” . Ding Ning and Ma Haoqin (2013) believe that scientific research data services are services provided by researchers and scientific research institutions for scientific research workers, mainly including scientific research data management planning services, data description and archiving services, scientific research data management education and training services, and related environmental services. Chen Yuanyuan and Ke Ping(2017)believe that research data services typically include creating and managing institutional data, providing data mining and visualization tools, training data management activities for researchers, providing institutional policy guidance, helping to develop data management plans and dataset metadata, and assisting in the resolution of intellectual property and privacy issues about research-related data., etc. In general, research data services are an activity carried out around the management and use of research data, and its content mainly includes research data management, research data analysis and data research literacy training. Regarding research data management services, most of them are based on the data management life cycle model theory, which is generally studied from the stages of data collection, data storage, data organization, data sharing, data analysis, and data reuse. The core services involved include needs identification of research data service, data management plan, metadata service, institutional knowledge base construction, data literacy and skills training. Research data analysis services mainly use data analysis tools to analyze research data and provide relevant analysis reports for researchers. Research data literacy training is also an important content of scientific research data services, mainly through online training, seminars, lectures and other means to enable researchers to understand the relevant knowledge of research data management, and enhance awareness and management capabilities about data management.

3.2 Job capability requirements for research data services

Based on the work content of research data services, some requirements are put forward for the job responsibilities and capability structure of data management positions. Some scholars have also conducted related research. Zhou Xiaoyan and Yin Yali (2016) analyzed the recruitment information of research data service personnel in the IASSIST website in 2015 and found that having strong data analysis ability and computer technology ability is one of the main conditions for the demand of research data service personnel in foreign university libraries. Cui Tao and Xu Jianguo (2017) believe that scientific data management talents should have three capabilities: basic literacy (including information, data, technology, etc.), knowledge structure (including computer skills, data management theory, management knowledge and other related subject knowledge) and intelligent structure (including thinking ability, organizational management ability and learning innovation ability, etc.). Through the interpretation and analysis of the Research Data Librarian Capability Framework, Sun Yuling (2017) believes that the core competency of research librarians' professional competence includes three modules: the ability to provide access to data, the advocacy and support capabilities of data management and the ability to manage data sets. Wei Lai and Zheng Huamin (2018) conducted a statistical analysis of recruitment documents about 170 foreign university libraries and 51 domestic universities libraries. They believed that data librarians should be based on data management, and their core competencies mainly include four aspects: 1) Data governance: metadata standards, data mining, data reuse; 2) data analysis: the use of data statistics software, the presentation of data visualization maps, common programming languages; 3) data sharing: reference consulting services, Open access, data life cycle; 4) Data management planning: project management experience, cooperation with research teams, language and writing skills. In short, research data service personnel are comprehensive composite talents. Only those who have both computer science and technology, social sciences, statistics and other professional knowledge, rich practical experience and comprehensive quality can be qualified for the position of research data services.

3.3 Problems in Research Data Services in colleges and universities

The research data management policies of foreign universities are mainly promoted by the state and research funding agencies/publishing institutions. In order to respect the relevant information and regulations formulated by the state and better strive for the research funding of research funding institutions, the universities have formulated coordination with all parties and consistent research data management policy. The policy content is comprehensive and detailed, focusing on the responsibilities and consultation and interpretation of the relevant stakeholders in the implementation process to ensure the enforceability and enforceability of the policy. China's university research data management and sharing policies are lacking or not mature, and the content is not comprehensive and systematic, and is still in the exploration stage.

At present, foreign universities have established research data service websites, and domestic universities have not yet appeared. Foreign universities have used the website to carry out research and education services for scientific research personnel with rich content and various forms of scientific research data, which have been developed in a specialized, precise and deep direction. China's scientific research data management education and consulting service practice started late, combined with the lack of scientific research data service website as a carrier, the data reference consulting service is single, the university does not train the scientific research personnel data management, and most of the university's research data management education It is a basic introduction to the management of scientific research data in universities. It does not cover the specific content, technical and software tools of the various aspects of scientific data management. It lacks the design of specific paths, and lacks the organic connection with the scientific research process of scientific research personnel, which seriously restricts the research data. The development and promotion of services.

The development of scientific research data services in foreign universities has established multi-field cooperation with various internal and external institutions. The construction of scientific research data service cooperation mechanism in foreign universities has played a powerful engine role in the improvement of scientific research data service level in universities. The level of cooperation in scientific research data services in China's universities is shallow. There is no mature coordination and cooperation mechanism between universities and institutions outside the university. The scope of cooperation is limited to the construction of scientific research data management platforms. The cooperation targets are limited to academic and academic institutions inside and outside the school. To a large extent, it will hinder the breadth and depth of the development of scientific research data services in China's universities.

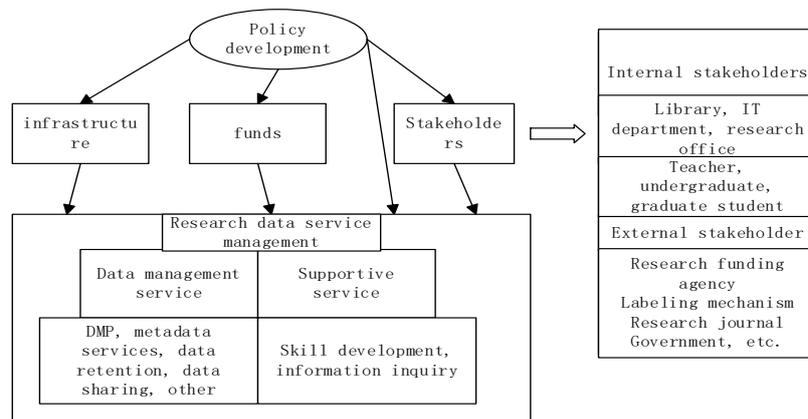


Figure. 1 "Double-class" university research data service model

4. Problems in the Cultivation of Data Talents in Universities of China

With the inclusion of “Big Data” in national strategic planning, China has also accelerated the pace of data science education. During 2014-2015, many well-known universities in China (such as Tsinghua University, Peking University, Xi'an Jiaotong University, etc.) took the lead in setting up the “Big Data Research Institute” and trained the first batch of graduate students in big data. Various universities and higher vocational colleges across the country have begun to research and declare big data majors around the construction of big data. By 2018, 283 colleges and universities have been awarded the “Data Science and Big Data” major. Data science major and data talent training has become one of the hot issues in China's current higher education. However, judging from the current training programs of some universities in China, there are some problems in the training of data talents.

At present, the domestic data science professional curriculum system is mainly composed of the main courses in the field of computer science and statistics. In the actual teaching, it is often only the original course teaching, and the two courses are not deeply integrated and applied. It has insufficient attention to basic theory of data science, data processing, data analysis, data calculation, data management, and data product development, and the data science major courses lack features. In addition, the curriculum design of data science majors in some domestic schools only emphasizes technical and engineering issues, while ignoring humanities and management issues, such as data policies, data ethics, laws involved in data, and cultural environments.

The lack of scientific research and innovation capacity training. Innovation is the driving force and source of scientific progress. As a college of higher learning that places scientific research at the forefront of running a school, research universities should attach great importance to the guidance and cultivation of students' scientific research innovation consciousness. However, in reality, the effect is very small, highlighting Lack of innovation capacity development. On the whole, college students lack interest and enthusiasm for research, and thus lack the motivation to study or conduct scientific research. The students' independent thinking is vague and lacks the scientific spirit of criticism and creation. Compared with foreign top university students, the lack of participation in innovative research activities such as research topics and academic exchanges, as well as the lack of high-quality innovative scientific research results, reflect the lack of scientific research and innovation capabilities of Chinese college students.

There is a lack of talent training and industry-university research. Under the current talent training mechanism, most research universities have embarked on the road of combining production, study and research, satisfying the technical needs in production through scientific research, and solving the technical problems encountered in real production. However, at present, the cooperation between industry, university and research institutes in domestic universities is still relatively weak. The technical needs of the school for the market are not very clear, so there is a deviation between the students and the talents in the development of students and talents. At the same time, there is no agreement between enterprises and universities on issues such as fund coordination and resource integration. The fragile cooperative

relationship makes students unable to obtain sufficient production internship opportunities, and it is difficult to grasp the professional practice skills required to enter the job, and it is difficult to be competent. The demand for talents in social, political, and economic development during the period.

Table 1 Evaluation indicators of scientific research data resource allocation in colleges and universities

University department	Research data resource output indicator		Scientific research data resource input index	
Earn rewards Y1	On Wenfa Expression Condition Y2	New main Project Y3	High level Research talent X1	Research project investment X2 (unit: 10,000 yuan)
Department awards	The number of the first/communication authors in the department's teachers	New teachers in the department Number of hosted projects	Number of professors, associate professors, lecturers, engineers, economics, and researchers in the department	Funding for research projects in the department

The degree of internationalization is low. International competitiveness and international influence are the true portrayal of the comprehensive strength of a university. Looking at the current domestic key research universities (211, 985 colleges and universities), although there are very few colleges and universities such as Peking University and Tsinghua University, the level is close to the world first-class, but in general, there is still a big gap with the world-class standards. In order to narrow this gap, the state has introduced corresponding policies to encourage exchanges and cooperation between domestic universities and overseas universities, especially the world-famous top universities. However, this relationship is still in its infancy and participates in exchange activities. The students are still a minority, and the cooperation between universities has not yet reached the ideal effect of driving the rapid development of domestic universities and catching up with the world first-class.

5. The training path of research data service talents for “Double first-class” construction

5.1 Strengthen professional education and set up a curriculum system connected to the market

Faced with the serious shortage of data talents, China has also accelerated the pace of data science education. Professional education is the main path for the cultivation of current scientific data talents. At present, China can train data management professionals by setting up a master's degree in data management, conducting certification projects, and offering courses. However, from the current situation of education, data talent education is mainly concentrated in the postgraduate education stage, and the output of data talents is far from meeting the needs of social development. With the development of society, the university community should reform the traditional talent training system and subject education system, call for the establishment of a professional degree in data management, set up a bachelor's, master's and doctoral degrees in "data science", and establish multi-type and multi-level scientific data management talents training system, which is conducive to establishing a stable mechanism for the output of data talents.

From the perspective of data processing needs in the context of big data, Ma Haiqun and Pu Pan (2016) divide data talent into seven types: data technology talents, data management talents, data security talents, data analysis talents, data policy talents, and data open talents. And data scientists. Colleges and universities must take the talent demand as the starting point, continuously strengthen the relevant professional curriculum of scientific data management, and cultivate specialized data talents according to the needs of the post work, and improve the scientific data management professional education curriculum system. Under the big data environment, traditional and specialized discipline talents have been difficult to adapt to the needs of social development. Scientific data talents must have some professional skills, but also master the theoretical knowledge of data management, which can be covered and satisfied the entire field of scientific data management from the level and structure. Therefore, different types of data talents should be cultivated in a focused manner, while considering the practical ability development of various types of talents, and setting up some applied courses and interdisciplinary courses. For example, in the curriculum setting, colleges and universities should open more compulsory or core courses for data management basic data preservation classes, data analysis classes, and metadata classes. These courses lay the foundation for students to master the core professional basic knowledge. Other compulsory courses include information organization and acquisition, digital service projects in the cultural heritage department, advanced project management, information analysis topics, database systems, big data, machine learning, information network applications, and technology communications. These courses consider the current information environment and will Information technology, information processing methods, information processing tools, etc. are combined with data management. (Wan Yan Deng, 2016)

5.2 Actively carry out multi-agency cooperation to achieve a combination of production, study and research

The data subject is a highly practical discipline. Its research and application cannot be separated from specific fields, and must also involve many factors such as development strategy, infrastructure, human resources, policy, legal and cultural environment. Therefore, data talent training is an interdisciplinary and cross-disciplinary social system engineering. In order to cultivate application-oriented data talents that meet the needs of society, colleges and universities must attach great importance to the cultivation of students' practical skills and professional abilities. In addition to the curriculum, colleges and universities should strengthen multi-agency cooperation, extensively use many forms of training, for example, internships, experimental classes, team work, practical projects, case studies and so on, to achieve the training of interdisciplinary, wide-calibre, applied data management professionals. The mechanism of cooperative training is diversified, such as joint between universities, joint between universities and research institutions, and joint between schools and enterprises. Extensive cooperation can provide students with a data environment and practical opportunities, which will help improve the business application capabilities and comprehensive practical capabilities of data service talents. For example, the University of Edinburgh and the UK's leading supercomputing center EPCC jointly develop data science high performance computing talent. At present, major Internet companies have also given various support to the training of talents in the field of data science. In June 2015, Alibaba Cloud and Huike Education Group launched the "Aliyun University cooperation program" (AUCP) to jointly open the core of cloud computing and data science. Through the integration of production and education, Alibaba Cloud provides resources such as cloud computing, big data engineers and ecosystem partners in AUCP to train the next generation of data scientists. In August 2015, ARM announced the establishment of "Guizhou University – ARM Innovation and Talent Training Base" in cooperation with Guizhou University to promote the development of local big data education and to cultivate big data talents for long-term sustainable development. Colleges and universities should further strengthen cooperation with enterprises, encourage more powerful enterprises to participate in the training of data talents, and provide training bases for building data talents.

5.3 Data literacy training for graduate students with professional disciplines

Carlson of Purdue University believes that "data literacy should include the following core competencies: understanding the format of databases and data, discovering and collecting data, data management and organization, data transformation and interoperability, data quality identification, meta Use of data, data preservation and reuse, data practices, data preservation, data analysis, data visualization, data ethics and citation specifications (Liu Caizhen, 2019). Data literacy is critical for researchers who need to be data quality scientists and potential data management professionals. Foreign countries attach great importance to data literacy skills training. Many colleges and universities have already produced data management training materials, which are embedded in the core curriculum of Dr. Shuo to enhance the graduate students' data literacy skills, so that researchers can be more closely connected with students and improve graduate students. The overall

research culture of the institute. For example, the Australian National University offers a comprehensive data management manual; the University of Northumbria incorporates data management training into the core skills curriculum of the PhD student training center and conducts annual data literacy training for interdisciplinary students. The US Health Sciences Library Data Services team developed materials for research data management and data visualization, and has provided courses by department, research group, and training program. The data service team invited teachers from the medical center with data expertise in the library. Teaching in a series of courses, and holding a series of eight brand courses called “Data Day to Day”, effectively solved the data skills gap of academic medical centers. The University of Minnesota conducts data management skills training for structural engineering graduate research teams and teacher advisors. Relatively speaking, domestic universities have basically ignored the data literacy of master doctors in professional disciplines. China needs to learn from foreign training experience extensively, improve the data literacy awareness of Dr. Shuo and improve its data management capabilities.

6. Conclusion

Under the background of “double first-class” construction, the value of scientific research data has become more and more prominent, the transformation of scientific research paradigm and the development of open access movement have given birth to the needs of open data and data management. The world’s major universities have participated in this trend. At the same time, it also provides opportunities for the cultivation of innovative and applied talents. Higher education institutions are better suited to the transformation and development of social needs. Through the cultivation of ideas, training channels and changes in training methods, it is also the core content of the new era society's demand for talents. In order to adapt to the new situation, we can effectively cultivate the innovative and entrepreneurial talents that the society needs, and then implement the strategic goals and tasks of “double first-class” construction. Based on the education and research big data platform, we will improve the teaching level and innovation ability of colleges and universities, and build an integrated talent training guarantee system with perfect content, sound standards, scientific operation, strong guarantee and remarkable results.

Acknowledgements

The authors acknowledge the Guangdong Philosophy and Social Science Planning Project: Research on the Construction of Research Data Service System of University Libraries under the Background of “Double-first Class Construction” (Project No.: GD17XTS06);Huizhou Science and Technology Project: Evidence Research and Development Strategy of Research Data Service in University Libraries(Project No.: 2017ZX116).

References

- [1] Huang Xin, Deng Zhonghua(2016). Research and Development Trend of Data-intensive Scientific Communication. Digital Library Forum, no. 5, pp.8-11
- [2] Cox, Andrew M.; Verbaan, Eddy. How academic librarians, IT staff, and research administrators perceive and relate to research. LIBRARY & INFORMATION SCIENCE RESEARCH, vol. 38, no.04, pp. 319-326
- [3] Ding Ning, Ma Haoqin(2013). Comparative Study and Reference of Scientific Data Life Cycle Management Models in Foreign Universities. Library and Information Service, vol. 57, no.06, pp. 18-22.
- [4] Chen Yuanyuan, Ke Ping(2017). Summary of Research on Scientific Research Data Service in University Libraries. Library Work and Study, no.10, pp.17-23+3
- [5] Zhou Xiaoyan, Yin Yali(2016). Analysis of the Knowledge Structure of Scientific Research Service Personnel in Foreign University Libraries--Taking the Recruitment Information of 2015 in IASSIST Website as an Example. Library and Information Service, vol.60, no.03, pp.76-82.
- [6] Cui Tao, Xu Jianguo(2017). Research on Models of Scientific Data Management Talents Cultivation from the Perspective of Library . Library Construction, no.03, pp. 84-89.
- [7] Sun Xiaoling(2017). The Construction Strategy of Data Librarians' Professional Ability: An Interpretation and Enlightenment Based on the Framework of Research Data Librarian's Capability . New Century Library, no.10, pp. 32-36.
- [8] Wei Lai(2018), Zheng Huamin. A Comparative Study of the Ability Requirements of Data Librarians at Home and Abroad. Library and Information Service, vol. 62, no.10, pp. 18-24.
- [9] Ma Haiqun, Pu Pan(2016). Reflections on the Cultivation of Data Talents in China from the Perspective of Big Data. Digital Library Forum, no.01, pp. 2-9.
- [10] Wan Yan Deng Deng, Gao Feng(2016). Investigation and Enlightenment of Foreign Data Management Talents Training. Library Theory and Practice, no.06, pp. 21-26.
- [11] Liu Caiwei(2019). Data Literacy: A New Content of Quality Education in Colleges and Universities . Journal of Beijing University of Technology (Social Science Edition), vol. 19, no.03, pp. 94-100.