The Development of Data Science Education in China from the LIS Perspective

Jilong Zhang, Anna Fu, Hao Wang, and Shenqin Yin

Abstract:

The aim of this paper is to introduce the development of data science in higher education in China, including the policy and educational programs at various levels. We investigated the data science education of five LIS (Library and Information Studies) schools in China, using Fudan University's Data Management and Application Master's Program as an example for more specific information about the curriculum structure, course focus and teaching methods in data science education. The paper further describes the action of promoting data science and data science education in the field of LIS by the China Academic Library Research Data Management Implementation Group.

To cite this article:

To submit your article to this journal:

Go to http://ojs.calaijol.org/index.php/ijol/about/submissions

The Development of Data Science Education in China from the LIS Perspective

Jilong Zhang, Anna Fu, Hao Wang, and Shenqin Yin

Fudan University Library, Shanghai, China

ABSTRACT

The aim of this paper is to introduce the development of data science in higher education in China, including the policy and educational programs at various levels. We investigated the data science education of five LIS (Library and Information Studies) schools in China, using Fudan University's Data Management and Application Master's Program as an example for more specific information about the curriculum structure, course focus and teaching methods in data science education. The paper further describes the action of promoting data science and data science education in the field of LIS by the China Academic Library Research Data Management Implementation Group.

Keywords: data science education, LIS programs, education of LIS, curriculum

INTRODUCTION

The era of big data has arrived with enormous challenges to every aspect of society. From science to education, healthcare, government, and commerce, the future of every sector is linked to the relatively new term "data science." People with specialized data skills are in urgent demand everywhere and a shortage of data talent appears in many countries (Manyika et al., 2011; Department for Business, Innovation & Skills, 2013; Liu & Jia, 2015). In 2016, Professor Yongwei Wu from Tsinghua University predicted that in three to five years China could face a shortage of as many as 1,500,000 data specialists (Bai, 2016). As a significant source of talent, institutions of higher education are considered a vital place for cultivating data scientists and specialists (Chen & He, 2016). In response, an increasing number of degree programs and courses in data science have been developed by universities in China.

As the disciplines that educate people who are capable of addressing big data challenges, data science and data science education have attracted intense attention from a wide range of domains and disciplines (Song & Zhu, 2016), The current paper aims to depict the state of data science education in China, especially from the LIS perspective.

LITERATURE REVIEW

Data science is closely related to statistics and computer science, and university students in these areas are usually seen as a primary source of potential data talent. However, faculties in these fields point out numerous opportunities and challenges they are facing when developing educational programs, and have provided an array of recommendations on curriculum design and educator development (Baumer, 2017; Ramamurthy et al, 2016, Bauman et al, 2014). Sun and Yin (2017) state that to cultivate data scientists with big data abilities, statistics programs need to: add courses related to the theory and software of big data, foster students' abilities in programing and computing, and integrate machine learning and big data analysis into their course content. However, Zhang and Huang (2014), suggested that Bayesian theory and neural network content be incorporated into teaching content, and that instruction should focus on students' practical and presentation skills and their sense of curiosity. They also highlighted the importance for statistics faculties to collaborate with faculties in mathematics and computer science. Regarding computer science education, Wang et al. (2017) identified data scientists, big data system engineers, big data algorithm engineers, machine learning engineers and big data algorithm scientists as five distinct data talents, and suggested enhancing the big data abilities of computer science students by developing course architecture with a focus on big data and big data tools. In addition to these two disciplines, with the significant impact of the new science paradigm, educational programs in journalism, economic management, business, publishing science, biological sciences and social science also become more data-centric with more concentration on their students' awareness and abilities to tackle big data problems (Shen et al.., 2014; Kirkpatrick, 2015; Wang & Liu, 2016; Bichler et al., 2017; Yu, 2017; Macmillan, 2015, Stephenson & Caravello, 2007).

As an academic field or academic program, data science is relatively new. However, there has already been a number of data science degree programs launched by universities, and schools or departments offering such programs include Business, Computer Science, Statistics, Mathematics, LIS, and Arts and Sciences (Tang & Sae-Lim, 2016). The variety of disciplines indicates the multidisciplinary nature of data science. Song and Zhu (2016) surveyed 48 data science programs in the US and found the number of Master's programs outnumbers certificate, Bachelor's and Doctoral programs. Most of these programs are developed by collaborations between multiple schools and departments. Research by Aasheim et al. (2015) and Tang & Sae-Lim (2016), found similar results and speculated that undergraduate programs may become a new trend in this area. The 2016 Park City Mathematics Institute (PCMI) organized a workshop and 25 faculties from computer science, statistics, and mathematics gathered to discuss their vision for data science education at the undergraduate level, and the appropriate content that should be covered by an undergraduate data science program. As a product of the workshop, a guideline for curriculum design of undergraduate programs in data science has been published (Veaux et al., 2017).

In the LIS field, an increasing number of job positions list data knowledge and skills as crucial requirements for the successful candidate (Chen & Zhang, 2017). Job titles are varied, and include data specialist, data curator, data librarian, data archivist, data scientist, and some other titles including the word "data." In response to increased demand, LIS schools start to revamp their programs by offering new data science and data curation programs and incorporating data science, data management and data curation courses into their existing programs (Tonta, 2016). Several authors surveyed the data science education offered by LIS schools and focused on the level of

education, curriculum structure, teaching content and approaches. Harris-Pierce & Liu (2012) noted that 16 LIS schools in the US and Canada offer courses on data curation and recommended more LIS schools add data curation courses to their programs. Si et al. (2013) investigated 63 scientific data courses offered by 25 iSchools around the world and found that the curriculum covered basic knowledge and methods of data curation, but lacked content such as data curation tools and user training approaches. Cao et al. (2016) surveyed 16 institutes offering data curation education listed by DCC website and nine of which were LIS Schools. Among these programs, fifty percent of core courses are in the area of LIS, such as information organization and access, database and information system, metadata, digital library, while the rest cover digital preservation and data curation technology.

Some case studies discuss the data science and data literacy education of LIS schools in more details. Within the context of e-Science, Qin and D'lgnazio (2010, 2016) introduced their course design and teaching experience from the Scientific Data Literacy project. Heidorn et al. (2007) described their experience of developing a Data Curation Education Program (DCEP) at the University of Illinois at Urbana -Champaign (UIUC). Huang and Ji (2015) suggested that LIS schools in China learn from the success of DCEP by focusing on data curation education, developing a project team, and organizing a relevant conference. Although some researchers claim that "data" has triggered next wave of curriculum changes in LIS schools (Tonta, 2016), Ma and Pu (2016) expressed great concern that compared to the Computer Science, Software Engineering and Statistics disciplines (which are the primary sources of IT and data talent), LIS schools still have a long way to go. After analyzing the course content of various programs in data science, Tang & Sae-Lim (2016) point out that iSchools' courses cover less content on mathematics and statistics than other schools and departments, and therefore, doubt the competitiveness of data science programs offered by iSchools.

DISCIPLINE CONSTRUCTION OF DATA SCIENCE IN CHINA

In order to produce appropriately skilled data talent, the Ministry of Education (MOE) of the People's Republic of China has taken a series of positive actions. In terms of undergraduate education, MOE identified data science, big data technology, information and computing sciences, statistics, applied statistics, computer science and technology, software engineering, information management and information systems as the seven most relevant disciplines to the area of big data (Ministry of Education of the People's Republic of China, 2016). In China, there are currently 2,638 higher education sites offering undergraduate programs in these fields. The breakdown is as follows: 983 in Computer Science and Technology, 563 in Software Engineering, 237 in Statistics, 107 in Applied Statistics, 526 in Information and Computer Science, 219 in Information Management and Information System, and 3 in Data Science and Big Data Technology (Ministry of Education of the People's Republic of China, 2016). Among these majors, Data Science and Big Data Technology are relatively new, as they were established by the MOE in 2015. The first three universities approved to offer Data Science and Big Data Technology programs are Peking University, University of International Business and Economics and Central South University. A year later, MOE approved requests from 32 additional universities to develop Bachelor's degrees in Data Science and Big Data Technology. Graduates of the data science programs from four of these universities obtain Bachelor of Science degrees, while the rest obtain Bachelor of Engineering degrees (Ministry of Education of the People's Republic of China, 2017a).

Since 2012, the MOE began releasing lists of undergraduate programs with low employment rates, in order to review and optimize the major structure of undergraduate programs in China. During this process, data science and big data technology became two of the prioritized majors and received strong support from the MOE. Universities were encouraged to improve their existing programs and to develop new programs relating to big data, based on the needs of social and economic development as well as the universities' existing teaching conditions (Ministry of Education of the People's Republic of China, 2017b).

Regarding postgraduate education, according to the guidelines in *"Regulation of Establishment and Management in Degree Granting and Talents-Nurturing Discipline Catalogue"* released by the China Academic Degree & Graduate Education Development Centre and MOE, second-level disciplines and specialties to be offered at the postgraduate level should be deliberated on by the universities themselves. Some well-positioned universities received support to develop programs in big data and cultivate graduates with the skills and competencies to handle big data and solve big data problems (Ministry of education of the people's republic of china, 2016).

Depending on the existing faculties in the departments or schools of information, computer science, and statistics, some Chinese universities have launched various data science courses and programs to bridge the gap between the required and current level of data scientists. For instance, Fudan University (FDU) launched a PhD program in data science in 2010 and two years later, they developed a postgraduate course titled Data Science. In 2015, FDU began to offer Master's programs in data science and their undergraduates could select data science as a minor. In 2013, Beihang University developed postgraduate programs in data science. In 2014, Tsinghua University announced the establishment of the Academy of Data Science and launched a number of multidisciplinary master's programs in data science. These programs are a collaboration between six faculties (e.g. Information School, Economic and Commercial School et al.) and are guided by the Postgraduate School. Programs such as data science, engineering and commercial analysis are built as the forerunners. In 2008, the Chinese University of Hong Kong developed a Master's program in data science and business statistics since 2008. Additionally, Chinese Academy of Sciences, Sun Yat-sen University and East China Normal University all have established their own research institutes to cultivate postgraduates in data science.

For vocational education, MOE released a revised version of the Catalogue Of University Diplomatic Education Majors (Ministry of Education of the People's Republic of China, 2015) in October 2015, taking the initiative in meeting the needs of the big data era, by adding new programs in cloud computing and application, technology of e-commerce and cyber data analysis and application. In 2016, MOE granted approval to 50 vocational colleges to launch cloud computing and application programs, and approval to 53 additional vocational colleges to begin offering technology of e-commerce programs (Ministry of Education of the People's Republic of China, 2016).

DATA SCIENCE EDUCATION OF LIS PROGRAMS

Investigation of Data Science Education in Five LIS Schools in China

Method

Five top LIS schools in China (Table 1) were used as a sample population to gain a general understanding of data science education in China's LIS schools. With the exception of the National Science Library of Chinese Academy of Sciences, which offers only postgraduate programs, the other four schools offer Bachelors, Master's and Doctoral programs.

Table 1.

Five sample LIS schools

University	LIS School	Website	Abbreviation
Wuhan University	School of Information Management	http://sim.whu.edu.cn/	WHU
Nanjing University	School of Information Management	http://im.nju.edu.cn/	NJU
Renmin University of China	School of Information Resources Management	http://irm.ruc.edu.cn/	RUC
Peking University	Department of Information Management	http://www.im.pku.edu.cn/portal/view.do	PKU
Chinese Academy of Sciences	National Science Library	http://www.las.ac.cn/	CAS

Documents including program descriptions, course catalogues, curriculum structures, course syllabi and school news on the websites of the sample LIS schools were searched and analyzed. When course information for a specific LIS school could not be found on its website, the authors browsed the websites of the undergraduate college or graduate school of the sample universities. Although an attempt was made to collect as much as information about these five LIS schools as possible, it was not possible to find some detailed or latest information about programs and courses on the websites.

Findings

The length of program description of each LIS school varies, but the basic information such as the name of the program is usually available. Our investigation indicates that LIS schools in China seem to be more confident in using information rather than data to describe their programs. As a result, the word "data" rarely appears in the name of the programs offered by the sample LIS schools. Even though, sometimes "data" can be found in the curriculum and course content. For instance, in addition to traditional LIS courses such as information organization and information description, the undergraduate program in library science at WHU involves courses such as database principle and application, and SPSS and data management were seen in experimental and practical teaching. The other typical LIS undergraduate program at WHU is information management and information system, courses of which include data structure and database system principle. Similar curriculum structure can be seen in the program descriptions for PKU and NJU. Both undergraduate programs of PKU cover courses such as statistics, data analysis, data mining, data structure and algorithm, while programs of NJU include data structure within core courses and data repository and data mining in other courses. Additionally, the master's program in library science at NJU offers database technology and the information science Master's program offers data mining technology and cloud computing.

Comparatively, data science education of LIS programs at RUC tends to be more systematic and its program description is more explicit in terms of data science education. One of its Bachelor's degree programs, Information Resource Management makes a clear statement that students in this program are educated to master the basic theory of information resource management as well as the theory and technology of data management. Graduates from this program are expected to be qualified for the positions of data organization, data locating, data processing, data storage, data retrieval, data disposal and data mining in companies, enterprises or government agencies (School of Information Resources Management of RUC, 2016). Courses in this program are more data-centric, and include data management, basis of data analysis, data repository, data mining, web analysis and text mining, social network analysis, Internet information analysis, data preservation, database system and data structure. The requirements for the information resource management Master's program are higher than the undergraduate program. The curriculum contains data mining and commercial intelligence, forecasting and deciding, competitive intelligence analysis, knowledge management, data repository, technology of data mining, and information forecasting. Furthermore, both Information Analysis Master's program and Library and Information Science Master's program at RUC provide students with courses in data mining and commercial intelligence, and data, model and decision. In terms of the intelligence theory and method Master's program, data mining and analysis are also included in the research content. The digital libraries program emphasizes database creation and management, data storage, data interchange format, data analysis and management technology, and data mining.

Guided by MOE, data science education in LIS programs at the graduate level in China seem to be more flexible and innovative than that at the undergraduate level. More elements of "data" occur in the description of the master's programs and PhD programs. One of the master's programs developed by CAS is titled "Methods and Technology of Big Data Intelligence Analysis." The program description is: "the international frontier technology including data mining, complex network, social network, text analysis and knowledge discovery, and particularly emphasizing the trend of research for the development, exploitation and technology of intelligence analysis, and

their application to the new methods, index and modes of informetrics." (National Science Library of CAS, 2017) By contrast, doctoral programs vary a lot due to the interdisciplinary character and wide application of data science, as well as the specialized nature of doctoral education. NJU offers a doctoral program named data intelligence and information system, while RUC offers the doctoral program in big data governance.

In addition to curriculum, on the web site of the School of Information Management of NJU, we found that they also encourage their students to be involved in extra-curricular activities and competitions relating to big data, data analysis and data application (School of Information Management of Nanjing University, 2015).

The Case Study of Master's Program of Data Management and Application at FDU

Introduction to the program

The Literature and Information Center (LIC) at FDU launched the master's programs in LIS in 2014, with the collaboration of the National Library of China, Shanghai Library, and Shanghai Science and Technology Intelligence Research Centre. At present, there are four two-year master's programs with six professors (three doctoral mentors included) and 27 associate professors. The Data Management and Application program provides a professional education and prepares students for data-intensive professional roles in a broad range of sectors. The program focuses on the knowledge and skills that LIS students require to manage data effectively in modern libraries and other information settings. Graduates are expected to be able to solve data management problems in modern libraries and other organizations with their knowledge in library science, information science, management science and computer science. The LIC has enrolled two batches of students since 2015, a total of 55 students, of which eleven selected data management and application program. These students have a wide range of undergraduate backgrounds from computer science, information management, information systems, applied mathematics and statistics to archive science.

What and how we teach

The Data Management and Application Program is a 24-month program and students are required to obtain 40 credits for graduation. Courses include: course learning, specialized practice, social practice, and a dissertation.

In terms of course learning, students are required to complete two introductory courses worth five credits, eight specialized courses for 23 credits, and four elective courses for seven credits. See Table 2 for detailed course information.

Table 2.

Course information

Course type	Course information	Teaching method	Faculty
Introductory Courses	Introduction to Library Science, Research Methods and Dissertation Preparation	Face-to-face	LIC
Specialized Courses	Development of Information Resources, Library Management, Information Service and User Research, Information Organization and Retrieval, Digital Library, Research and Implementation of Data Curation, Information Analysis and Visualization, The Frontier and Dynamics of LIS	Face-to-face	LIC
Elective Courses	Such as second language, Academic Writing in English	Face-to-face	Other schools, e.g. College of foreign language and literature,

The specialized curricula of the data management and application programs fall into two categories: traditional and typical LIS courses including the development of information resources, information service and users, information organization and retrieval, and the frontier and dynamics of LIS, and specialized data science courses including the technology of digital library, research and application of data curation and information analysis and visualization. All of the courses are important for preparing students for their future careers in the library and information profession. However, compared to the former category, the later category focuses on data policy, management, analysis and service. Although Digital Library remains the course title, the course content differs from the previous course. In addition to digital library technology, information systems and project management, the course incorporates the concepts and technology of big data, data management and services. The Research and Application of Data Curation course is designed to introduce research data management, and, from the data curation point of view, to provide students with knowledge and skills in data selection and evaluation, data ingestion, long-term data preservation and data access and dissemination. This course gives students a broad view of the background of research data management, the history of the development of data curation, the modules and tools of data curation, as well as national and international practices. Students are also required to master the theoretical knowledge of data curation, and apply it into daily research activities to effectively manage research data and make data management plans. The objective of the information analysis and visualization course is to foster students' abilities in data mining, data analysis and data visualization, with the help of computer tools including MS Excel, SPSS, R, Python, Tableau and Citespace.

Course instructors not only focus on knowledge acquisition, but also provide students with opportunities to understand the value of data, data related professions and work involving data by

inviting guest speakers. For instance, last semester, the digital library course invited a chief engineer working at a computer company, a manager from an Internet finance company, a specialist of a population data platform, and a data librarian from a US academic library to present to students. These four guest speakers with diverse work experience and educational backgrounds introduced what they do with data and what they think data specialists should be. This allows students to connect "data" with their learning and future careers. Additionally, interdisciplinary teamwork is crucial for preparing data specialists in today's data-driven world, so group work is used widely in this program. In the research and application of data curation course, students are divided into three groups at the beginning of the semester. One of the group assignments is to investigate the Data Management Plan (DMP) Online and DMP tool. Each group is required to choose a tool to study and explore its function and implementation. About two weeks later, groups present their findings with slides. Assignments also play an important role in the master's program, especially for courses such as information analysis and visualization. The instructors of these courses hope that their students use the appropriate tools to analyze and use data after their learning. Therefore, they have designed a series of assignments to help students become skilled in using tools instead of just knowing the tools.

In addition to the curriculum, students in LIC at FDU are required to have a 6-month internship during their two year education. LIC has established a few formal internship sites both domestic and abroad, including at: Migrated Population Service Centre of China National Health and Family Planning Commission, EastMoney.com, Shanghai Population Data Research Centre and a variety of libraries. During the internship, students work on real-world data problems in the specialty fields under the mentorship of multi-mentor teams from both LIC and their workplace. Practice content includes management, sharing and service of migrated population investigation data, disposal and application of financial big data, open government data, population data sharing, and the management and service of research data. Nearly fifty percent of students' dissertations are relevant to their internships and under collaborative guidance and supervision from both their program supervisor and internship mentors. Three of the graduates this year focus on demographic data, while the other two students studying the reliability of research data repository, and the exploitation and utilization of adolescent research data separately. All these topics are very relevant to their internships and the research projects they engaged in.

Reflections and Discussion

LIS faculty are always facing challenges in developing curricula that meet the needs of students and employers. When LIS meets "data," the nature and scope of LIS work changes a lot. It is impossible to solve the problem by waiting for LIS students to graduate with more data science skills; LIS faculty should be more creative and proactive in how they integrate new knowledge and skills into their curriculum. Three core courses concerning big data, data curation and service, and data analysis have been included within the data management and application program, but the faculty of FDU are still trying to make the current curriculum more data-centric. More courses geared towards developing students' statistical and computing skills need to be added in the future, so that the graduates can be more competitive compared to graduates from data science programs in other disciplines.

Diverse discipline backgrounds give our students varied standpoints and provide them with knowledge and specialized skills when studying and doing research. However, it is clear that such

diversity makes it more difficult for faculty to design curriculum and assignments, especially from a teaching perspective. To provide students with different training based on their undergraduate backgrounds, interests and future career plans, faculty need to pay more attention to integrating specific projects into the program.

Further, skills and abilities required by academic libraries are different from healthcare, new media or Internet, and the latest technology may soon become obsolete. It is impossible to cover every aspect of data science in a two year degree program, and to produce graduates with complete knowledge and skills in data science. Therefore, the LIS faculty at FDU is planning to include contents such as life-long learning and open access materials relevant to data science into the students' learning process.

PROMOTING DATA SCIENCE IN THE LIS FIELD

The results of the survey report on MLIS (Master of Library and Information Studies) program development in China (Wang, 2015) indicate that data mining has become one of the most useful courses in LIS programs. Survey respondents suggest that the teaching content in statistic tools, programming languages and data mining in LIS programs needs to be expanded and enriched. It is obvious that data science has become a crucial component of LIS education. How should LIS practitioners promote the awareness of data science and what can be done by LIS faculties to improve the educational content and pedagogical structure? Such urgent questions need to be discussed in the LIS field. In this context, the China Academic Library Research Data Management Implementation Group (CALRDMIG) was established by nine university libraries, including FDU Library, PKU Library, Tsinghua University Library, Shanghai Jiao Tong University Library, Zhejiang University Library, Beijing Institute of Technology Library, Shanghai International Studies University Library, Tongji University Library and Wuhan University Library in 2014. This section uses CALRDMIG as an example, to introduce the efforts of LIS practitioners to promote data science and related education.

To have a crucial role in the future development of the data science ecosystem, LIS professionals need to know very well about the changes, potential and challenges of data to the development of LIS. Therefore, in the first annual meeting, every member of CALRDMIG selected specific topics in relation to research data management based on their organizational strength and strategic priorities. Topics included but were not limited to the following: scanning environment, organizing the framework of managing research data, drafting the measures and policies of managing research data, drafting and implementing the standards and regulations of research data, investigating the platforms and tools, selecting the model for the system, localizing software and system in China and secondary development, training sessions and programs for managing research data, drafting the developing plan of research data management service for university libraries and providing best practice, and building supportive environment for using and applying research data. Members meet every year to discuss the new progress of each area.

Data science education as a topic of broad and current interest has been discussed many times. For instance, at the 2015 annual conference at the Beijing Institute of Technology, participants from nine university libraries discussed the education and training of managing research data. Three aspects were covered: data literacy education for library users, data science education for LIS students, and post education and professional training for librarians. In addition

to library managers and senior librarians, young librarians and data librarians were also invited to join the meeting, where participants could share their experiences and challenges they meet in a data-intensive context.

Compared to irregularly held conferences with a loose pattern of communication, it is much more effective to build a cohesive community of enthusiastic data science LIS practitioners, faculties and researchers. Three years into its development, CALRDMIG has generated some stimulating outcomes. Members can share their experience of data management, data service and cultivation of data talents, with LIS people and people from organizations in other fields and sectors. The 2016 annual conference was held at the Shanghai University of Foreign Studies on May 18. After that, CALRDMIG members organized a forum named "Shaping Intelligence, Gaining Value from Data: China Academic Research Data Management & Information technology for Library (IT4L) Conference, 2016" with a total of 120 participants from more than 60 university and public libraries. Apart from data professionals in the LIS area, Professor Xizhe Peng from FDU's Social Science Data Research Centre, Deputy Director Weidong Wang from RUC's Social Investigation Data Centre, and Associate Professor Deqing Yang from FDU's School of Big Data gave speeches to the forum. At the end of conference, the Group released a proposal for universities and research institutes in China to: raise awareness of data management and data sharing, promote the drafting of related policies, communicate and collaborate with the relevant fields to research data, and boost technological and research innovation by developing data collecting, archiving, publishing, sharing, and using.

CONCLUSION

Institutions of higher education play an important role in helping develop data talents. In China, with the influence of a series of measures from the MOE, an increasing number of programs and institutes in data science have been established. To keep pace with the transformation of big data, it is incumbent on LIS faculties to ensure that their curricula are effective and appropriate in preparing graduates with competencies for data-intensive workplaces. The results of this investigation indicate that LIS schools in China have integrated data science into their degree programs from undergraduate to postgraduate level, although different faculties focus on different knowledge and skills in the area of data science. Data science education at the postgraduate level is more systematic and explicit. However, in contrast to LIS programs in the US, the UK and other western countries, LIS faculties in China use data to describe their programs much less. The authors of this paper believe that data science education in Chinese LIS programs is still in the early stage of development.

The Data Management and Application Master's program has been implemented by LIC at FDU for more than two years. Although it is not possible to cover all data science knowledge and skills within a two year master's program, graduates of the program are expected to be competitive and have appropriate skills to cope with data problems. In the future, more courses on statistical and computing skills, especially the latest data technologies will be integrated into the program. Given the multidisciplinary nature of data science, more collaborations with other departments, such as School of Data Science at FDU may be necessary. MOOCs may act as supplementary courses for solving the problem of time constraints in the master's program.

Furthermore, LIS faculty at FDU will continue to focus on students' ability to solve real-world problems and try to provide them with more real-world project opportunities.

It is time for LIS faculties and researchers to do more research on how data science education should be better incorporated into LIS programs in China. A number of Chinese researchers have surveyed the educational programs of data science and data curation in North America and Europe and provided suggestions on learning experience from those existing educational programs. However, few specifically investigate and concentrate on the data science education offered by Chinese LIS schools and faculties. This study investigated the five top LIS schools in China using the material available on their web sites. This paper introduced a case study of FDU, and contributes to a basic understanding of the current state of data science education in LIS programs in China. Further research could include more LIS schools and materials to gain a better understanding of the development and effectiveness of data science education provided by LIS programs in China.

References

Aasheim, C. L., Williams, S., Rutner, P., & Gardiner, A. (2015). Data analytics vs. data science: a study of similarities and differences in undergraduate programs based on course descriptions. *Journal of Information Systems Education*, *26*(2), 103-115.

Bai, H. (2016). *Expert predicted: The shortage of data talent in China is about a million and a half.* Retrieved from http://news.cyol.com/content/2016-05/30/content_12687384.htm. [in Chinese]

Bauman, P. T., Chandola, V., Patra, A., Jones, M. (2014). Development of a computational and data-enabled science and engineering Ph.D. program. *2014 Workshop on Education for High Performance Computing (EduHPC)*, New Orleans, Louisiana. November 16-21, 2014. Piscataway, NJ, USA: IEEE Press. doi:10.1109/EduHPC.2014.8

Baumer, B. (2015). A data science course for undergraduates: Thinking with data. *American Statistician*, *69*(4), 334-342.

Bichler, M., Heinzl, A., & Aalst, W. M. P. V. D. (2017). Business analytics and data science: Once again? *Business & Information Systems Engineering*, *59*(2),1-3. doi: https://doi.org/10.1007/s12599-016-0461-1

Cao, R., Wang, Q., Geng, Q., & Liu, X. (2016). Investigation and analysis of data curation education in foreign universities. *Library and Information Service*, (4), 52-58. doi:10.13266/j.issn.0252-3116.2016.04.007. [in Chinese]

Chen, H. & Zhang, Y., (2017). Educating data management professionals: A content analysis of job descriptions. *The Journal of Academic Librarianship*, *43*(1), 18-24. doi:http://dx.doi.org/10.1016/j.acalib.2016.11.002

Chen, Z., & He, T. (2016). Data science: The demand and development of talents. *Big data research,* 5, 95-106. [in Chinese]

China Academic Degree & Graduate Education Development Centre & Ministry of Education of the People's Republic of China. (2009). *Regulation of establishment and management in degree*

granting and talents-nurturing discipline catalogue. Retrieved May 6, 2017, from
http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_834/200903/45419.html. [in
Chinese]

Department for Business, Innovation & Skills. (2013). *Seizing the data opportunity: A strategy
for UK data capability*. Retrieved from
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-
1250-strategy-for-uk-data-capability-v4.pdf

Harris-Pierce, R. L., & Liu, Y. Q. (2012). Is data curation education at library and information
science schools in North America adequate? *New Library World*, *113*(11/12), 598-613.

Heidorn, P., Palmer, C., Cragin, M., & Smith, L. (2007). Data curation education and biological
information specialists. In DigCCurr 2007: *An international symposium in digital curation*.
Chapel Hill, NC, April 18-20, 2007. Retrieved from
http://www.ils.unc.edu/digccurr2007/papers/heidornEtal_paper_8-2.pdf

Huang, R & J, C. (2015). Current situation and enlightenment of the data curation education in
University of Illinois at Urbana-Champaign. *Library & Information*, *1*, 61-65.
doi:10.11968/tsygb.1003-6938.2015010. [in Chinese]

Kirkpatrick, K. (2015). Putting the data science into journalism. *Communications of the ACM,
58*(5), 15-17. doi: 10.1145/2742484

Liu, B. & Jia, Y. (2015). Cultivating talents in the age of big data in Japan: Background, research
direction, plan and measures. *E-Government*, (10), 85-95. doi:
10.16582/j.cnki.dzzw.2015.10.012. [in Chinese]

Ma, H. & Pu, P. (2016). The pondering of cultivation for talents in data under the circumstance
of big data in China. *Digital Library Forum*, (1), 2-9. [in Chinese]

Macmillan, D. (2015). Developing data literacy competencies to enhance faculty collaborations.
Liber Quarterly the Journal of European Research Libraries, *24*(3), 140-160. doi:
10.18352/lq.9868

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., & Roxburgh, C., & Byers, A. H.
(2011). *Big data: The next frontier for innovation, competition, and productivity.* Retrieved from
http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-
frontier-for-innovation

Ministry of Education of the People's Republic of China. (2017a). *Notice to the 2016 approved
majors at undergraduate level by the Ministry of Education of the People's Republic of China*.
Retrieved May 6, 2017, from
http://www.moe.edu.cn/srcsite/A08/moe_1034/s4930/201703/t20170317_299960.html. [in
Chinese]

Ministry of Education of the People's Republic of China. (2017b). *Accelerate the transformation
of administration and service in higher education*. Retrieved May 6, 2017, from
http://www.moe.gov.cn/jyb_xwfb/s271/201704/t20170406_301996.html. [in Chinese]

Ministry of Education of the People's Republic of China (2016) *Ministry of Education of the
People's Republic of China's reply to the advice NO.2174 of the fourth session of the 12th NPC*.

Retrieved May 6, 2017, from
http://www.moe.edu.cn/jyb_xxgk/xxgk_jyta/jyta_gaojiaosi/201611/t20161102_287366.html. [in Chinese]

Ministry of Education of the People's Republic of China. (2015). *Catalogue Of university diplomatic education majors (2015).* Retrieved May 6, 2017, from:
http://www.moe.edu.cn/srcsite/A07/moe_953/201511/t20151105_217877.html. [in Chinese]

National Science Library of China Academy of Sciences (2017). *The description of postgraduate programs in LIS of National Science Library, China Academy of Sciences.* Retrieved May 6, 2017, from http://www.las.cas.cn/yjsjy/pyyxw/pyfa/201704/U020170406457750783870.pdf. [in Chinese]

Qin, J., & D'Ignazio, J. (2010). Lessons learned from a two-year experience in science data literacy education. *International Association of Scientific and Technological University Libraries, 31st Annual Conference*, 188–204. Retrieved from
http://docs.lib.purdue.edu/iatul2010/conf/day2/5

Qin, J. & D'Ignazio, J. (2016). Enhancing scientific data literacy in college students: Experience and lessons learned. *Journal of Librarianship & .Information Studies*, *8*(1), 1-27. doi: 10.6575/JILA.2016.88.01

Ramamurthy, B. (2016). A practical and sustainable model for learning and teaching data science. *SIGCSE '16 Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. Memphis, Tennessee, USA. March 2-5, 2016. New York, NY, USA: ACM. doi: 1145/2839509.2844603

School of Information Resources Management, Renmin University of China (2016). Undergraduate programs. Received from School of Information Resources Management, Renmin University of China website, Retrieved May 6, 2017, from
http://irm.ruc.edu.cn/more.php?cid=234. [in Chinese]

School of Information Management of Nanjing University (2015). Academic research and scientific innovation. Received from School of Information Management of Nanjing University website, Retrieved May 6, 2017, from http://im.nju.edu.cn/content.do?mid=6&mmid=66. [in Chinese]

Shen, H., Tan, H., & Wen, L. (2014). Data literacy of data journalists. *Youth Journalist* (21).17-18. doi: 10.15997/j.cnki.qnjz.2014.21.014. [in Chinese]

Si, L., Xing, W., Guo, W., & Zhuang, X. (2013). The cultivation of scientific data specialists. *Library Hi Tech*, *31*(4), 700-724. doi: https://doi.org/10.1108/LHT-06-2013-0070

Song, I., & Zhu, Y. (2016). Big data and data science: what should we teach? *Expert Systems*, *33*(4), 364-373. doi:10.1111/exsy.12130

Stephenson, E., & Caravello, P. S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review*, *35*(4), 525-540. doi: https://doi.org/10.1108/00907320710838354

Sun, X. & Yin, B. (2017). Statistics in the context of big data age. *Statistics & Decision*, *6*, 2+189. [in Chinese]

Tang, R & Sae-Lim, W. (2016). Data science programs in U.S. higher education: An exploratory content analysis of program description, curriculum structure, and course focus. *Education for Information*, *32*(3), 269-290

Tonta, Y. (2016). Developments in education for information: Will "data" trigger the next wave of curriculum changes in LIS schools? *Pakistan Journal of Information Management & Libraries*, *17*, 2-12. Retrieved from: http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tonta-pakistan-jiml-2016-888-2369-1-SM.pdf

Veaux, R., Agarwal, M., Averett, M., Baumer, B.S., Bray, A., Bressoud, T. C., …Tiruviluamala, N., Uhlig, P., Washington, T. M., Wesley, C. L., White, D., & Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application*, *4*(1), 15-30. Retrieved from: http://www.stat.berkeley.edu/~nolan/Papers/Data.Science.Guidelines.16.9.25.pdf

Wang, C. Y. (2015). *Survey report on MLIS program development in China* (Master thesis). Available from CNKI Master's Theses database. Retrieved from http://xueshu.baidu.com/s?wd=paperuri%3A%28481f34b67597895200f1fd0c92abdfc6%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Fcdmd.cnki.com.cn%2FArticle%2FCDMD-10269-1016126721.htm&ie=utf-8&sc_us=2266641113642335464. [in Chinese]

Wang, H., Gao, H., Yin, S., & Zhu, J. (2017). The design of course architecture for big data, Proceedings of the ACM Turing 50th Celebration Conference- China (ACM TUR-C '17). Shanghai, China. May 12-14, 2017. New York, NY: ACM.

Wang, X, & Liu, H. (2016). Importance of application of data science in experimental teaching of data science in experimental teaching of economics and management specialty. *Experimental Technology and Management*, *33*(4), 179-181. [in Chinese]

Yu, W. (2017). Editors' big data literacy. *China Publishing Journal*, (5), 12-16. [in Chinese]

Zhang, H & Huang, S. (2014).Approaches to develop big data analysis ability of students in Statistics. *Statistics and Decision*, *24*, 66-68. [in Chinese]

About the authors

Jilong Zhang is Associate University Librarian, Executive Deputy Director of the Institute for Humanities and Social Science Data, and Deputy Director of the Social Science Data Research Center at Fudan University.

Anna Fu is Data Librarian, Department of Data Management and Technical Service at Fudan University Library.

Hao Wang is Master's Student at the Literature and Information Center at Fudan University.

Shenqin Yin is Digital Library Projects Manager of the Department of Data Management and Technical Service, and Assistant Director of the Social Science Data Research Center at Fudan University.