A Study on the Current Status of Data Literacy among Chinese Graduate Students in Research Activities: A Dual Lifecycle Theory Perspective (Postprint)

Authors: Yu Weijie, Zhou Yali, Wu Jinchi

Date: 2023-04-01T16:15:52+00:00

Abstract

[Purpose/Significance] From the perspective of dual lifecycle theory, this study investigates the current status of data literacy among Chinese graduate students in research activities, and proposes recommendations for improving their data literacy based on the survey and analysis results. [Method/Process] Through a systematic review of domestic and international concepts and connotations of data literacy, an initial evaluation framework for data literacy was constructed. Building upon this, based on the embedded relationship between the data lifecycle and the research lifecycle, a graduate student data literacy evaluation framework under the dual lifecycle was further developed. Using the target skills in this framework as a reference, a questionnaire was designed and distributed to graduate students from different universities, disciplines, and academic years across China, and the survey results were analyzed. [Results/Conclusion] The survey results reveal that Chinese graduate students possess strong data ethics and legal awareness, but their data capabilities are generally weak, and deep-level data awareness remains to be enhanced. Data literacy exhibits variations across different disciplines and academic years. Meanwhile, corresponding solutions are proposed based on these findings.

Full Text

Preamble

Investigation and Research on the Status of Chinese Postgraduates' Data Literacy in Scientific Research Activities—From the Perspective of Dual Life Cycle Theory

Yu Weijie, Zhou Yali, Wu Jinchi School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [Purpose/Significance] Based on dual life cycle theory, this paper investigates the current status of data literacy among Chinese graduate students in scientific research activities and proposes recommendations for improvement. [Method/Process] By reviewing data literacy concepts and connotations domestically and internationally, a preliminary evaluation system for data literacy was constructed. Building upon this, according to the embedded relationship between data lifecycle and research lifecycle, a graduate student data literacy evaluation system based on dual lifecycles was further developed. Using the target skills in this system as a reference, a questionnaire was compiled and distributed to graduate students across different universities, majors, and grades nationwide, with results analyzed. [Result/Conclusion] The survey found that Chinese graduate students possess good data ethics and legal awareness, but their data capabilities are generally weak, and deeper-level data awareness needs improvement. Data literacy also varies across different disciplines and grades. Corresponding solutions are proposed based on these findings.

Keywords: data literacy; research lifecycle; data lifecycle; dual lifecycle

---

1 Research Status

Since M. Schield first proposed the concept of data literacy in his 2004 paper "Information literacy, statistical literacy, data literacy" [1], universities and related institutions at home and abroad have vigorously promoted data literacy education to adapt to the big data era. As professional scientific research talents, graduate students must develop data literacy as one of their core competencies.

Currently, through the efforts of numerous scholars, new teaching models for enhancing students' data literacy capabilities are gradually taking shape. Most domestic and international scholars have evaluated graduate students' data literacy status based on either the research lifecycle or data lifecycle alone, which provides guiding significance for promoting graduate data literacy work. However, periodic research on graduate data literacy development is essential. By establishing a graduate data literacy evaluation system under the dual lifecycle framework, this study explores the current status of graduate data literacy based on dual lifecycles, aiming to provide references and suggestions for the practice of graduate data literacy education.

In terms of the connotation of data literacy, scholars hold various perspectives. In 2007, E. Stephenson et al. pointed out that data literacy is an awareness and ability to effectively and appropriately discover, evaluate, and use information and data [4]. Foreign scholar K. Hogenboom et al. conducted comprehensive research on data literacy, defining it as a comprehensive ability to read, interpret, analyze, think critically about data, and use data as evidence [5]. P. J. Calzada believed that through programmatic design and self-evaluation, combined with characteristics of effective data sources, types, and acquisition channels, appropriate methods should be identified to manipulate, analyze, and quantitatively present data, with problem-solving approaches centered on data demonstrating data literacy capability [6]. In recent years, domestic experts have also conducted research on data literacy. In 2013, Ma Yunpeng, from the perspective of primary mathematics education, proposed that cultivating data literacy in contemporary primary students must emphasize data collection and statistical abilities, as well as the ability to question and reflect on data sources and formats [7]. In the same year, Zhang Jingbo analyzed the skills required in research processes involving data acquisition, organization, and analysis, arguing that researchers should adhere to basic ethical norms when processing data and achieve effective data use throughout the data lifecycle [8].

In 2015, the Association of College & Research Libraries (ACRL) passed the Framework for Information Literacy for Higher Education in the United States [9], which integrates data literacy as part of the academic process while emphasizing that data literacy education should maintain periodicity and embeddedness even when integrated into academic activity lifecycles, as data literacy education has its own lifecycle. In 2016, the University of Virginia Library conducted a practice on "Data Literacy Competencies Embedded in Research Workflow and Data Lifecycle," guided by the data lifecycle model embedded in research workflow on its official website. Specifically, the library website used research lifecycle navigation to provide specialized data information services for users at different research stages [10]. Cornell University Library, King's College London Library, and the University of Sydney Library have also followed this model to provide researchers with services throughout the entire research lifecycle, including high-level research consultation, data analysis, academic communication, intellectual property, and publishing services [11].

In recent years, domestic scholars have also introduced lifecycle theory into data literacy research. Hao Yuanling et al. proposed that the rise of data-intensive research paradigms has made it increasingly difficult for university researchers to collect, manage, publish, and cite scientific research data, thus requiring stricter capabilities in handling research data, particularly regarding attitudes toward data and data analysis skills [12]. Zhang Jun constructed a researcher data literacy competency cultivation framework based on research data lifecycle theory, using research projects as the main subject [13]. Hu Hui et al. sorted out the core content system of data literacy embedded in research workflow and data lifecycle, mapping out the data literacy competency framework that researchers should possess [3].

Currently, scholars mostly study data literacy in combination with a single lifecycle. However, both research lifecycle and data lifecycle are cycles aimed at data operation and value realization, being both unified and intersecting. From this perspective, combining data literacy research with dual lifecycles can enrich research dimensions and enable more precise, comprehensive, and scientific investigation of data literacy. Therefore, establishing a data literacy evaluation system based on dual lifecycles provides a new perspective for exploring the status and causes of differences in graduate data literacy. Based on these findings, suggestions can be made for graduate data literacy education, holding practical significance for improving graduate data literacy levels.

2 Research Design

This study examines data literacy among Chinese graduate students in scientific research activities, including graduate students from different universities, majors, and grades. The survey method was employed, combining the dual lifecycle model to compile a questionnaire on graduate data literacy based on dual lifecycles. Both online and offline distribution were used, with scientific allocation according to actual master-doctoral student ratios.

The research approach involves: (1) introducing domestic and international research findings on dual lifecycles and data literacy; (2) combining characteristics of domestic graduate student groups to compile a questionnaire on their current data literacy status under the dual lifecycle framework; (3) analyzing the data literacy situation of graduate student groups from both overall and dual lifecycle perspectives; and (4) proposing recommendations for improving graduate data literacy.

3 Model and Evaluation System Construction

3.1 Dual Lifecycle Model

In 2006, scholar J. Humphrey proposed a knowledge transfer model consisting of six stages: conceptualization, initialization, analysis, initial results generation, formalization, and promotion, which laid the foundation for the research lifecycle. Subsequently, many scholars studied the research lifecycle. The UK Joint Information Systems Committee believed the research lifecycle should contain five stages: concept formation, collaboration seeking, writing, research, and publication [14]. Li Wenwen divided the research lifecycle into four phases: research initiation, research planning, research implementation, and results publication [15]. North Carolina State University Library divided the research lifecycle into five parts: idea generation, funding seeking, proposal, ongoing research, and dissemination, providing corresponding support services for researchers according to the characteristics of each period [11]. Referring to the research stages that graduate students experience, this study prefers to divide the research lifecycle into four phases: research initiation, research preparation, research implementation, and results publication.

Regarding data lifecycle models, official authoritative institutions and scholars have proposed various structures. The UK Data Archive established a circular lifecycle diagram for researchers and data-intensive groups, detailing the data competencies required at each stage from "research planning - data collection -

data processing and analysis - data publication and sharing - data preservation - data reuse" [16]. The UK Joint Information Systems Committee established a closed-loop data lifecycle cycle for university researchers and research management support personnel, defining the data lifecycle as "from data creation and storage - data management - databases and archives - data catalogs and registration - planning and design - collection and acquisition - collaboration and analysis - management and storage - sharing and publication - discovery and reuse" [17]. The International Standards Association's Data Documentation Initiative Alliance released the DDI (Data Documentation Initiative) data lifecycle model, which differs from other institutional models by adopting a semi-circular structure, believing the data lifecycle should include "concept - collection - processing - distribution - discovery - analysis - reuse - processing - archiving - distribution" steps [18]. Compared with other models, the UK Data Archive (UKDA) model is concise, comprehensive, and universal. In terms of presentation, this study believes that graduate students' daily data usage processes align more closely with the semi-circular lifecycle proposed by DDI, as not every component needs to participate in the lifecycle cycle. Using these two lifecycles as references, this study constructed a data lifecycle model (see Figure 1 [Figure 1: see original paper]).

In graduate students' research lives, the research lifecycle and data lifecycle are inseparably related. Scientific research itself is an exploration of data, embedding substantial data work in research activities, while the accumulation of research data further promotes research development. The two are closely connected and mutually reinforcing. Based on this, dual models of research lifecycle and data lifecycle have gradually become a research trend, with scholars conducting in-depth studies from perspectives of demand relationships [16], content commonalities [19], and integration patterns [20].

As research on the dual lifecycle concept is still in its infancy, this study conducted a comprehensive analysis and integration of current domestic and international research on data lifecycle and research lifecycle models, combining relevant research practices to ultimately construct a dual lifecycle model framework (see Figure 1).

The dual lifecycle in Figure 1 is built upon the data lifecycle and research lifecycle, with different boxes representing different lifecycle stages while reflecting how the data lifecycle is embedded in the research lifecycle to explain the relationship between the two cycles.

In the research lifecycle, the research initiation phase is the beginning of research, dominated by awareness and involving simple data collection capabilities. After establishing the research direction in the initiation phase, students enter the research preparation phase, which involves collecting and analyzing large amounts of data to prepare for research implementation. This phase mainly corresponds to data collection and processing stages in the data lifecycle. The research implementation phase serves as a connecting stage, primarily including research experiments and paper writing. Research experiments organize data obtained from experiments and external sources, process them through various means to reach research conclusions, and finally present these conclusions through visualization or textual forms. The results publication phase requires researchers to publish data in journals and databases corresponding to their research fields, thus corresponding to the data publication and preservation stages at the end of the data lifecycle. Data preservation includes not only the final published data but also the organization and preservation of unpublished data obtained during experiments.

3.2 Data Literacy Evaluation System

As researchers deepen their study of data literacy, the construction of data literacy evaluation systems has become a focus of exploration. E. S. Gummer et al. constructed a data literacy evaluation framework based on domain analysis, with evaluation indicators including problem identification, question formulation, data use, data-to-information conversion, information-to-decision conversion, and result evaluation [21]. S. Z. Athanases et al. built a teacher data literacy evaluation system based on the education sector, with indicators including data operation ability, problem articulation ability, data collection and organization ability, data analysis tool usage ability, and data presentation ability [22]. These evaluation systems focus on data capabilities, while Cao Shujin et al. argued that data literacy should emphasize not only data capabilities but also data awareness and data ethics [5].

In summary, this study believes that among the three components of data literacy, data awareness determines vision and height in research activities, data capability determines the depth and breadth of research content, and data ethics determines behavioral norms during the research process.

After reviewing relevant research and referencing the 2015 U.S. authoritative "Framework for Information Literacy for Higher Education" [24], combined with the characteristics of data lifecycle and research lifecycle and the actual conditions of graduate research activities, this study established a data literacy evaluation system (see Figure 2 [Figure 2: see original paper]).

The first-level indicators in this system consist of data awareness, data capability, and data ethics—the three components of data literacy discussed above. Data awareness includes three second-level indicators: acquisition awareness, usage awareness, and dissemination awareness. Data capability includes four second-level indicators: data collection ability, data processing and analysis ability, data presentation ability, and data preservation ability. Data ethics includes three second-level indicators: data norms, data security, and data law. Under these ten second-level indicators, 19 target skills are further defined.

3.3 Data Literacy Evaluation System Based on Dual Lifecycle Theory

Current data literacy education can be summarized from horizontal and vertical perspectives. Vertically, data literacy education progresses step-by-step from data awareness, data capability, to data ethics, gradually advancing from awareness cultivation to capabilities in data collection, application, and storage. This vertical approach stands from the perspective of scientific data resources, using the data lifecycle as the main thread for education from resource navigation, general education, to disciplinary literacy. Horizontally, data literacy education examines implementation from the research lifecycle perspective, from research initiation to results publication.

As mentioned earlier, the dual lifecycle is a model with the research lifecycle as the main thread and the data lifecycle embedded within it. Based on the dual lifecycle model in Figure 1 and the graduate data literacy evaluation system in Figure 2, combined with the characteristics of data literacy needs at each stage, this study constructed a data literacy evaluation system based on dual lifecycles (see Figure 3 [Figure 3: see original paper]).

This evaluation system based on dual lifecycles not only reflects the embedding relationship between the two lifecycles but also uses different numbers to correspond to target skills in the data literacy evaluation system (for example, number 1 corresponds to the first target skill "Recognizing the importance of data" in Figure 2), more intuitively demonstrating the data literacy skill requirements at different stages across lifecycles. This evaluation system serves as the core for questionnaire design, with surveys conducted to explore current graduate data literacy.

4 Analysis of Graduate Students' Data Literacy Status

4.1 Survey Design

The survey subjects were graduate students, selected through random sampling. To ensure that data literacy levels were not affected by geographical factors, the survey included graduate students from northern China (Beijing, Heilongjiang), central China (Chongqing, Sichuan), and southern China (Guangdong). To avoid bias due to institutional differences, the survey included samples from "985" universities, "211" universities, and ordinary institutions. The survey covered both doctoral and master's students, with master's students further categorized by grade and by professional vs. academic type. Graduate disciplines included humanities and social sciences, natural sciences, engineering and technical sciences, medicine, agriculture, and other fields. Based on a five-point Likert scale, the questionnaire scoring mode increased from left to right, with scores ranging from 1 to 5. To present data more intuitively, graduate data literacy levels were assessed using percentages: 85% and above as excellent, 70%-85% as good, 60%-70% as qualified, and below 60% as unqualified.

4.2 Reliability and Validity Analysis

A total of 754 questionnaires were collected from graduate students in different regions. After removing 4 incomplete or obviously random responses, 750 valid questionnaires were obtained, with a recovery rate of 99.47%. Using SPSS software for reliability and validity analysis (excluding demographic questions), the Cronbach's Alpha coefficient was 0.862, indicating very good reliability. Bartlett's test of sphericity yielded a coefficient of 0, and when this coefficient is less than 0.05, the questionnaire has structural validity. Therefore, this questionnaire possesses structural validity.

4.3 Overall Analysis of Graduate Data Literacy

As constructed in the data literacy evaluation system, questionnaire items were designed for each indicator. The data awareness section included 5 items covering acquisition awareness, usage awareness, and dissemination awareness. Data capability, as the most important component, included 10 items covering data collection, processing and analysis, presentation, and preservation abilities. The data ethics section included 4 items covering data law, data security, and data norms. Valid questionnaires were compiled and average scores calculated (see Table 1).

Table 1 shows that graduate students' overall data awareness is above average. Specifically, most students clearly understand the importance of data for research, which received the highest score in data awareness. Students also performed well in taking responsibility for published data and ensuring data fairness and openness. However, performance was weaker in deeper awareness aspects such as critical thinking about data and data sensitivity and insight, reaching only qualified levels. This indicates that graduate data awareness remains at relatively superficial levels, with weaker awareness for deep data mining and critical thinking.

Data capability is the core component of data literacy education. As shown in Table 1, among all data capabilities, data preservation ability scored highest, followed by data presentation ability, then data processing and analysis ability, with data collection ability scoring lowest. Overall, graduate students' data capabilities remain relatively weak. Particularly concerning is the lowest-scoring data collection ability—specifically, the ability to scientifically organize and code collected data barely reached the passing level. Good scientific organization and coding capabilities can make subsequent research more efficient, making this an essential capability that cannot be ignored in graduate data literacy education. Graduate students also showed weak abilities in deep data mining, a highly valuable new field that is technically difficult and rarely widely applied in daily research activities.

Data ethics are fundamental principles everyone must follow in data activities. Table 1 shows that most graduate students comply with data laws and ethics, can use data securely, and properly cite sources when publishing data. However, they performed relatively weaker in data format and type standardization compared to other data ethics aspects. This suggests that while students pay attention to red-line issues like data law and security, they are less attentive to data standardization.

4.4 Dual Lifecycle-Based Analysis of Graduate Data Literacy

This study designed the questionnaire with the research lifecycle as the main thread. Based on the dual lifecycle-based graduate data literacy competency framework constructed earlier, the study analyzed graduate data literacy status from both research and data lifecycle perspectives. Exploring graduate data literacy under dual lifecycles is significant for better cultivating data awareness and management capabilities.

Figure 4 [Figure 4: see original paper] presents the data literacy status derived from the dual lifecycle model. The left table shows research lifecycle stages with corresponding data literacy target skills, while the right table shows data lifecycle stages with their corresponding skills. Dotted lines connecting the two tables more intuitively represent the correspondence of target skills across dual lifecycles and their embedding relationship.

The left side of Figure 4 shows graduate data literacy status across research lifecycle phases based on scores and corresponding evaluation levels. Data literacy scores were highest during research initiation, followed by results publication, both rated as "good." Research implementation and preparation phases were rated as "qualified." Across the entire research lifecycle, students scored highest in research initiation and lowest in research preparation. The research preparation phase, where researchers collect and integrate materials according to research purposes, significantly relates to data processing and conclusion derivation in the research implementation phase.

The right side of Figure 4 shows graduate data literacy status across data lifecycle phases. Students achieved "good" comprehensive target skill levels in the research topic formulation and data publication phases, but only "qualified" levels in data collection, processing, and preservation phases, with data collection scoring lowest and data processing second lowest. The figure reveals that data processing and collection phases are data capability-intensive stages, while the higher-scoring topic formulation and data publication phases are less data capability-intensive. In graduate students' daily research, data collection and processing are critical stages determining research success.

4.5 Comparison of Graduate Data Literacy Based on Dual Lifecycle

4.5.1 Positive Correlation Between Data Literacy Level and Grade

To explore data literacy status across different grades, graduate students were categorized into four groups: first-year master's, second-year master's, third-year master's, and doctoral students. Based on the dual lifecycle stage division, data literacy status was compared across grades at each lifecycle stage to explore the relationship between grade level and data literacy (see Table 2 , where scores are averages of corresponding items at each stage).

In the research lifecycle, significant differences exist across grades. From first-year to third-year master's students, data literacy levels at each research lifecycle stage show an increasing trend. In research initiation and results publication phases, doctoral students' data literacy is significantly higher than third-year master's students, while in research preparation and implementation phases, differences are not significant, with doctoral students even scoring slightly lower than third-year master's students.

In the data lifecycle, the positive correlation between data literacy level and grade is more evident. Overall data literacy levels of doctoral and third-year master's students are similar; second-year master's students outperform first-year students at all stages; third-year master's students scored lowest in the topic formulation stage. Except for third-year master's students scoring lower than first and second-year students in topic formulation, and doctoral students scoring lower than third-year master's students in data processing, graduate data literacy level generally shows positive correlation with grade level.

4.5.2 Significant Disciplinary Differences in Data Literacy Level Due to obvious differences in knowledge attributes, research paradigms, disciplinary cultures, and output forms across disciplines, all disciplines were divided into six categories: humanities and social sciences, natural sciences, engineering and technical sciences, medicine, agriculture, and others, to compare data literacy levels among graduate students across disciplines (see Table 3 , where scores are averages of corresponding items at each stage).

In the research lifecycle, natural science students ranked highest overall, scoring highest in research implementation and results publication phases, followed by medicine and agriculture students, while "other disciplines" scored lowest. In the research lifecycle, science and engineering graduate students scored higher than humanities students, which relates to their daily research environment and demands.

In the data lifecycle, natural science and agriculture students ranked highest overall. Comparing differences across disciplines at each data lifecycle stage, the largest variation occurred in the topic formulation stage, with a range of 0.52 across disciplines. The smallest variation occurred in the data processing stage, with a range of 0.26, showing the most balanced scores.

These results demonstrate significant disciplinary differences in graduate data literacy levels. Specifically, science and engineering students with stronger technical capabilities show higher data literacy levels than humanities students with less data capability demand. Similarly, students in data-intensive disciplines show higher levels than those in disciplines with less data usage. These gaps may relate to graduate students' data literacy levels in different majors or to their inherent data capability status.

5 Recommendations for Improving Graduate Data Literacy

5.1 Strengthen Dual Lifecycle-Based Graduate Data Literacy Education

Domestic data literacy education can reference foreign content design, using the research lifecycle as the main thread. Horizontally, education should cover data collection, organization, management, preservation, utilization, and reuse, as well as training in related policies, data management, data sharing, and data ethics. Vertically, through general data literacy education, a progressive educational format should emphasize hierarchical advancement in educational models [25]. As libraries play important roles in university data literacy education, university libraries will provide more training for graduate students, including data query and data sharing.

In daily university teaching, progressive education models should be developed according to data literacy characteristics to support improvement under the dual lifecycle framework. General data literacy education should be conducted around research workflows. In terms of awareness, sensitivity to data needs, data security, and data ethics should be enhanced to guide students in establishing correct data values and 善于利用数据提高专业技能 [26]. From a usage perspective, practical courses should be established to 普及 data management knowledge such as data resource management, construction, and data support services, enabling students to master basic data usage skills.

5.2 Enrich Dual Lifecycle-Based Data Literacy Education Forms

Currently, China's university data literacy education forms and content settings are relatively 单一, mainly consisting of library lectures supplemented by training and specialized education, with some universities offering MOOC courses as sub-courses of information literacy. Drawing from foreign experience, the University of Illinois at Urbana-Champaign, for example, offers multi-domain data management education courses covering information organization and ethics [27].

Survey results show that the current 短板 of Chinese graduate students' data literacy lies in data capability. Future data literacy education in China could form research groups where students experience research workflows to identify data management problems and solutions; universities could provide diverse data capability elective courses for students to choose according to their majors and needs; and offer low-credit mini-courses for self-selected stages in the lifecycle requiring data literacy improvement.

5.3 Promote Multi-Level Graduate Data Literacy Education

Different grades and majors show varying data literacy status and needs. Therefore, tailored, multi-level data literacy education should be implemented for different graduate types.

The positive correlation between grade level and data literacy indicates that research experience duration and knowledge accumulation are the fundamental causes of differences across grades. As newcomers to research, first-year master's students should focus on enhancing data awareness and ethics education, while senior students engaged in intensive research should emphasize data capability education in collection and processing. The survey found that the gap between doctoral students and third-year master's students is not significant, which does not align with the data literacy doctoral students should possess. Future data literacy education for doctoral students should simultaneously strengthen data awareness, capability, and ethics education with greater depth and breadth than master's students.

Due to differences in research field nature and knowledge structure, students from different majors exhibit distinct data behaviors in research activities, as evidenced by the disciplinary differences in data literacy status. Science students generally outperform humanities students, and data-intensive disciplines outperform less data-intensive ones. Drawing from UCLA's discipline-embedded teaching collaboration model, subject instructors and librarians should cooperate in data literacy education [5].

Overall, graduate students currently show low critical thinking and sensitivity toward data, with deeper data awareness needing improvement; data capabilities are generally weak and require strengthening; but they possess good data ethics and legal awareness. From the dual lifecycle perspective, graduate students perform well in research initiation and results publication phases but poorly in research preparation and implementation phases. In the data lifecycle, students score highest in research topic formulation but only achieve qualified levels in data capability-intensive collection, processing, and preservation phases, indicating considerable room for improvement.

This study aims to provide references for data literacy education implementers and researchers, and offers a new perspective for cultivating and improving national graduate data literacy levels.

References

[1] SCHIELD M. Information literacy, statistical literacy and data literacy [J]. IASSIST quarterly/International Association for Social Science Information Service and Technology, 2004, 28(2): 7-. [2] Song Junyu. Research on university library data literacy service platform based on "Internet +" [J]. New Century Library, 2017(8): 50-55. [3] Hu Hui, Wu Ming. Research on data literacy competencies embedded in research workflow and data lifecycle [J]. Library and Information, 2016(4): 125-137. [4] STEPHENSON E, CARAVELLO PS. Incorporating data literacy into undergraduate information literacy programs in the social sciences: a pilot project [J]. Reference services review, 2007, 35(4): 525-540. [5] Cao Shujin, Liu Huiyun, Wang Weijia. From information literacy to data literacy [J]. Library and Information Research,

2017(1): 19-24. [6] CALZADA PJ, MARZAL M. Incorporating data literacy into information literacy programs: core competencies and contents [J]. Libri: international journal of libraries and information services, 2013, 63(2): 123-134. [7] Ma Yunpeng. Connotation and value of core competencies in primary mathematics [J]. Primary Mathematics Education, 2015(9): 3-5. [8] Zhang Jingbo. Data literacy education in the big data era [J]. Science, 2013, 65(4): 29-32, 4. [9] Qin Xiaoyan. Improvement and enlightenment of American university information literacy standards—Interpretation of ACRL's "Framework for Information Literacy for Higher Education" [J]. Library and Information Work, 2015, 59(19): 139-144. [10] Research data management [BE/OL]. [2019-06-28]. http://guides.library.vcu.edu/data. [11] Xue Jingjing, Jiao Kun, Zhang Xin, et al. Research on foreign university library research support services based on research lifecycle [J]. Information Studies: Theory & Application, 2016, 39(5): 110-114. [12] Hao Yuanling, Shen Tingting. Construction and strategic thinking of data literacy and its cultivation mechanism [J]. Information Studies: Theory & Application, 2016, 39(1): 58-63. [13] Zhang Jun. Research on researcher data literacy cultivation facing the fourth research paradigm [J]. Library and Information, 2016(2): 133-136. [14] Suo Chuanjun, Wang Tao, Fu Guangyu. Review of information lifecycle management research at home and abroad [J]. Library Journal, 2008(7): 14-20. [15] Li Wenwen, Zhou Ping. Research on scientific literacy education based on research lifecycle [J]. Modern Intelligence, 2017, 37(1): 156-159. [16] UK DATA SERVICE. Research data lifecycle [EB/OL]. [2019-06-30]. https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx. [17] LYON L. Dealing with data: Roles, rights, responsibilities and relationships [EB/OL]. [2019-06-30]. https://purehost.bath.ac.uk/ws/files/419529/dealing_{{{with}}*{{data}}*}{report}-final.pdf. [18] Overview of the DDI version 3.0 conceptual model [EB/OL]. [2019-07-02]. http://opendatafoundation.org/ddi/srg/Papers/DDIModel_{v_4}.pdf. [19] Ling Wanyang. Research on researcher data literacy under big data and data-intensive research paradigm [J]. Library, 2018(1): 81-87. [20] Ding Ning, Ma Haoqin. Comparative study and reference of foreign university scientific data lifecycle management models [J]. Library and Information Work, 2013, 57(6): 18-22. [21] GUMMER ES, MANDINACH EB. Building a conceptual framework for data literacy [J]. Teachers college, 2015, 117(4): 1-22. [22] ATHANASES SZ, BENNETT LH, WAHLLEITHNER JM. Fostering data literacy through preservice teacher inquiry in English language arts [J]. The teacher educator, 2013, 48(1): 8-28. [23] Yao Xiaolu. Composition and cultivation of embedded subject librarian data literacy in the big data era [J]. Agricultural Library and Information Science Journal, 2017, 29(2): 198-201. [24] Association of College and Research Libraries. Framework for information literacy for higher education [EB/OL]. [2019-07-03]. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/Framework_{ILHE}.pdf. [25] Zhang Qun, Liu Yumin. Research on university library scientific data literacy education for graduate students [J]. Journal of Academic Libraries, 2017, 35(3): 93-98. [26] Ma Xiaohui. Discussion on college student data literacy education under big data background [J]. Software Engineering, 2019, 22(2): 48-50, 41. [27] Meng Xiangbao, Li Aiguo. Research on foreign university library scientific data literacy education [J]. Journal of Academic Libraries, 2014, 32(3): 11-16.

Author Contributions: Yu Weijie: Topic selection, research design, paper writing and revision; Zhou Yali: Data collection, initial draft writing; Wu Jinchi: Data processing and analysis.

Note: Figure translations are in progress. See original paper for figures.