# A Review of Domestic and Overseas Research on Data Literacy Education

Yuqing Tang[1] and Lu Han[2(✉)]

[1] School of Humanities, Beijing University of Technology, Beijing, China
[2] Beijing University of Technology Library, Beijing, China
`hanlu311@bit.edu.cn`

**Abstract.** With the education of data information literacy in the era of data-intensive scientific research as the main research objective, the thesis reviews the connotation and origin of data literacy education at home and abroad, the theoretical and practical research on carrying out data literacy education, etc. It analyzes the differences in data literacy education at home and abroad, and points out the problems of data literacy education in China including the lack of theoretical framework of curriculum, the lack of research on data literacy competency exploration, and the lack of education practice.

**Keywords:** Data literacy education · Data literacy · Information literacy

## 1 Introduction

The data-intensive research paradigm is the main driving force behind the rise of data literacy research, which has led to a wave of data-based scientific research, making data literacy a necessary competency for researchers in scientific research and a core competency for scientific innovation [1]. In early 2012, the Big Data Research and Development Initiative(BDRI) was launched [2]. Since then, data literacy has gradually become a hot topic of research for scholars at home and abroad. More and more scholars have been actively exploring data literacy from different aspects and have formed rich theoretical and practical achievements. This paper systematically reviews the research results in the field of data literacy from the perspective of data literacy education, in order to provide reference and inspiration for future research on data literacy education.

## 2 The Connotation of Data Literacy

Data literacy has not developed a universally accepted standard definition since its emergence. Foreign researchers have emphasized the ability to collect, organize, analyze, process, and evaluate data in a holistic manner [3]. Domestic researchers tend to describe data literacy in a hierarchical manner [4]. In addition to the above-mentioned holistic and hierarchical descriptions of data literacy application capabilities, some scholars focus on the formulation of data literacy from the perspective of comprehensive capabilities,

which is more general and abstract [5]. Some scholars focus on data literacy from the perspective of comprehensive competency, which is more general and abstract. Some other scholars explain the connotation of data literacy more specifically and purposefully from the scientific data utilization process, and they believe that data literacy emphasizes the ability to understand, use, and manage scientific data [6]. The definition of data literacy by domestic scholars is more focused on the specific hands-on ability of individuals to work with data, while foreign scholars place more emphasis on the higher-level ability to make decisions based on data analysis in terms of data-based operations, of which data awareness, data competence, and data ethics are the three most core dimensions. With the development of data literacy and related research, on the basis of the concept of data literacy, many related concepts have been gradually derived, such as data management literacy, scientific data literacy, data information literacy, etc. The above concepts are essentially similar to "data literacy".

## 3   The Origin of Data Literacy Education

The earliest foreign research on data literacy education dates back to 2004, when Love [5] early research on data literacy education was discussed, and the National Sanitation Foundation-funded "Using Data Project" was described in some detail to help math and science educators develop data literacy. By engaging teachers in a rigorous dialogue about data analysis and reflection, it will improve reach and learning in math and science and ultimately close the achievement gap. Research on data literacy education in China is relatively recent [7]. In his article "On the Development of Data Literacy in Physics Education", he first combined the subject of physics with data literacy in an effort to promote deeper integration of information technology and physics teaching by enhancing students' data literacy levels.

The rise of a data-intensive research paradigm, user needs, and active practices in the library community have driven the development and transformation of data literacy education. As early as 2012, the United States launched an innovation series of programs and initiatives focused on scientific data literacy education, namely the Data Information Literacy program, which is jointly funded by the American Institute of Museum and Library Services and aims to develop the ability of future researchers to locate, organize, process, and share data, which is an important contribution to the emergence and development of data literacy education. This has been an important contribution to the emergence and development of data literacy education [8]. Data literacy is rooted in information literacy. Data literacy is rooted in information literacy [4]. Data literacy is the ability to acquire, manage, share, and use data throughout the life of data in the era of big data, but it is not the same as information literacy.

## 4   Current Status of Data Literacy Education Research at Home and Abroad

IN recent years, research on data literacy education has been carried out at home and abroad, but the focus of their research differs greatly. Foreign theoretical research started

earlier and has formed some curriculum system and carried out some practice, while domestic is still at the stage of theoretical research and a few institution of higher education have carried out practice, but a systematic curriculum system has not yet been formed.

### 4.1 Theoretical Study of Data Literacy Education

#### 4.1.1 Research on the Content of Data Literacy Education

IN 2004, Hunt K discussed the probability of adding data literacy development to university information quality development; in 2017, Stephenson E et al. added data quality development to the examples of information quality development in social sciences [9]. Light states that data literacy indicates the ability to collect, classify, summarize, analyze, transform data into information, and assist in decision making. It also includes the ability to make or critique decisions based on data [10]. According to Prado's analysis, data literacy includes the generation and reuse of data, the identification of values, types and formats of data, and the ability to assess, curate, organize and self-evaluate data as important components [11]. The Data Quality Campaign consider data literacy as acquiring data, integrating and analyzing data, understanding other types of information data, using different types of data information in the teaching process to conduct periodic studies of data drivers, optimizing existing teaching strategies, and implementing personalized assessment to deepen students' understanding of data and to enable data communication. This will deepen students' understanding of data, enable data exchange and cooperation, and provide powerful guidance for teaching practice.

Chinese scholars Xiangbao Meng et al. [12]. Argue that in order to develop good data literacy, it is necessary to consider all aspects of data literacy in a comprehensive manner, including data awareness, data competence, and data ethics. Changliang Zhang et al. [13]. Proposed that the process of data literacy education in colleges and universities should be combined with the needs of social development, emphasize the importance of data literacy needs in professional fields, and strengthen the teaching of data processing theory and practical experience. On the basis of summarizing the research and practice of scientific data literacy education in university libraries at home and abroad, Qun Zhang [14] constructed the "5W" three-dimensional model of scientific data literacy education system in university libraries, including the participating subjects of scientific data literacy education, the intervention stage of scientific data literacy education in libraries, the driving factors of data literacy education in libraries, the content of data literacy education in libraries and the implementation methods.

#### 4.1.2 Establishment of Data Literacy Education System

Koltay [15] in his study focused on the importance of data literacy education in college libraries and argued that data literacy should establish a uniform terminology. Frank et al. [16] presented data information literacy education in college libraries using meteorological subjects as a case study. Dillo [17] focused on co-education of archival and library data management in the study, creating a front as well as back office model. May-bee et al. [18] used a rooted theory approach in the paper to examine the syllabus

of a nutrition science and political science course at Purdue University to analyze the information literacy and data literacy needs of students.

Libraries are the main institution for information literacy education in universities, based on which it is more beneficial to further develop data literacy education. Qun Zhang and Yumin Liu [14] constructed the "5W" three-dimensional model of university library data literacy education system, and it is pointed out that university libraries will be transformed from information centers to scientific data centers in the future. Xun Liu's [19] study points out that in order to effectively carry out data literacy education, university libraries in China must actively seek cooperation with multiple parties, such as professional faculty, embedded courses with online teaching platforms to achieve the purpose of education; cooperation with students, embedded in the student environment; and cooperation with researchers, embedded in the whole process of research projects. Chen Zhang [20] summarized the changing library service functions in the era of big data. Firstly, the library will become a social information center in the future, and with that, it will become a social library and continuous education.

## 4.2  Practical Research on Data Literacy Education

The most important way to cultivate data literacy is to design reasonable and effective curriculum and training programs [21, 22], and some related teaching practice activities have been carried out at home and abroad to offer data literacy course for students of different majors, and the specific contents offered by some universities are shown in Table 1. For example, Purdue University and Harvard University, which are more representative in the United States, carry out corresponding teaching programs and training in order to meet the needs of students' data literacy improvement. The Johns Hopkins University School of Medicine and Harvard Medical School in the United States offer data literacy courses to help medical students improve their ability to extract key information and draw scientific conclusions from the complex data of big data technology onion [23]. Some universities in China have combined general data skills with educational practices, such as Wuhan University to take advantage of the library literature data resources to carry out training in data analysis software, literature database utilization and other practical education [24]. Nanjing University focuses on the development skills of data literacy from the perspective of computer technology. Fudan University adds data literacy courses to the "Scientific Exploration and Technological Innovation" curriculum module for undergraduates, forming a cross-fertilization between courses, using the teaching method of social science data platform, introduction data spaces and data sets, and providing training on the use of Data verse to meet the basic data literacy needs of students. This course meets the basic data literacy needs of students. Most domestic institutions offer courses related to "data science and big data technology", most of which are based on data structure, database and other computer knowledge systems to set the teaching content, mainly for students of related majors, not suitable for students without relevant knowledge reserve [25], while the data literacy course offered by libraries mainly focus on the course of data literacy in libraries are mainly focused on the use of documentary databases and the use of related software. Currently, there is a lack of practice in building data literacy general education courses for specialized fields [26].

**Table 1.** Introduction to Data Literacy Courses and Programs in China and Abroad

| College or institutions | Course/program Name | Course Content |
|---|---|---|
| Purdue University | Data Information Literacy Program offers "Fundamentals of Data Thinking" course | Data awareness, data collection and acquisition, database management, data mining, data visualization, data ethics and morality |
| Harvard University | Data Interaction Community Project Conducts "Data Scientist" Training | Data management technology, data processing tools, data service infrastructure construction |
| Johns Hopkins University School of Medicine | "Clinical Informatics" and "Clinical Data analysis" course are offered. | Approaches to data management, data analysis and data visualization in the medical fields |
| Harvard Medical School | "Medical Data Decision Making" course series | Theoretical courses on statistical decision making, modern information science, machine learning, etc. |
| Wuhan University | Data Literacy and Data Utilization General Education Course | Data acquisition, data processing, data sharing, data distribution |
| Nanjing University | R language and data analysis, playing with data in Python | Teaching data storage, processing, and analysis using R or python programming language |
| Fudan University | Behind the Data, Information Literacy and Scientific Discovery, and Informed Thinking courses | Modules based on the "Scientific Exploration and Technological Innovation" curriculum |

## 5   Conclusions

(1)   The research on "data literacy" in China is later than that in foreign countries, and there are fewer theoretical achievements, and many of the theoretical achievements are based on the borrowing of foreign theoretical achievements. Foreign scholars mainly study the specific content of data literacy, how to cultivate students' data literacy, and how to conduct periodic research on data; domestic scholars mainly learn from the relevant foreign research, and study the connotation of data literacy, the importance of data literacy education, and the current situation of data literacy education.

(2)   In terms of the establishment of education system, foreign scholars integrate data literacy with specific disciplines and carry out different data literacy cultivation programs for students of different disciplines. And domestic scholars currently

mainly draw on the experience of foreign libraries, or analyze foreign university libraries as cases, and analyze the elements of data literacy education by comparing the characteristics and differences of data literacy education at home and abroad, so as to think about the future development direction of data literacy education in China's universities, but data literacy education needs to be in line with the country's situation.

(3) In terms of the practice of data literacy education, many foreign universities have already put data literacy education into practice, and more foreign universities have carried out some data literacy courses and programs, and achieved effective results. Moreover, with the extensive development of data literacy education practice, the audience of data literacy education has gradually expanded. There are fewer practices about data literacy education in China, and only some universities in China have carried out some data literacy education courses, such as embedded data literacy education and general data literacy education carried out by libraries, drawing on the experience of foreign countries.

# References

1. Zhang Qun, Liu Yumin. Research on scientific data literacy education in university libraries for graduate students[J]. Journal of University Libraries,2012,35(03):93-98.
2. LISA-JOY Z. PRESS RELEASE: Obama Administration Unveils "Big Data" Initiative: Announces \$200 Million in New R&D Investment [EB/OL]. [2020-06-09]. https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unvels-big-data-initiative-announces-200.
3. MAINACH E B, GUMMER. A Systemic View of Implementing Data Literacy in Educator Preparation [J]. Educational Research,2013,42(1).
4. Huang Ruhua, Li Baiyang. Data literacy education: the expansion of information literacy education in the ear of big data[J]. Knowledge of Library and Information, 2016 (01):21-29.
5. Love N. Taking data to new depths[J/OL]. Journal of Staff Development, 2004 [2020-06-01].
6. Qin J, D'LGNAZIO J. Lessons Learned from a Two-Year Experience in Science Data Literacy Education [EB/OL]. [2020-06-10]. http://docs.lib.purdue.edu/iatul2010/conf/day2/5/.
7. Xiang Hua. On the cultivation of data literacy in physics education[J]. China Modern Education Equipment, 2011(04): 84-85.
8. Hao, Yuanling, Shen, Tingting. The construction and strategic thinking of data literacy and its cultivation mechanism [J]. intelligence Theory and Practice,2016,39(01):58–63.
9. Stephenson E, Caravello P S. Incorporating data literacy into undergraduate information literacy programs in the social sciences: a pilot project [J]. Reference services review, 2007, 35(4).525–540.
10. Kim K S, Chiu M H, Sin S, et al. Recruiting a diverse workforce for academic/research librarianship: career decisions of subject specialists and librarians of color[J]. College & Research Libraries,2007,68(6):533-533.
11. Prado J C, MÁ Marzal. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents[J]. Libri,2013,63(2):123–134.
12. Zhang Xuan, Meng Xiangbao. A case study of data literacy education in universities for digital humanities[J]. journal of University Libraries,2019,37(05):87–94.
13. Zhang Changliang, Wang Chenxiao, Li Jingtong. A comparative study on data literacy education in Chinese and American universities in the era of big data[J]. Intelligence Theory and Practice,2019,42(08):131-137.

14. Zhang Qun, Liu Yumin. Research on the construction of a model of scientific data literacy education system in university libraries[J]. Journal of University Libraries,2016,34(01);96-102.
15. Tibor Koltay. Data literacy: in search of a name and identity[J]. Journal of Documentation,2015,71(2).
16. Emily P. Frank, Nils Pharo. Academic Librarians in Data Information Literacy Instruction: A Case Study in Meteorology.[J]. C&RL,2016,77(4):536–552.
17. Ingrid Dillo, René van Horik, Andrea Scharnhorst. Training in Data Curation as Service in a Federated Data Infrastructure - the Front Office-BackOffice Model.[J]. CoRR,2013, abs/1309.2788.
18. Clarence Maybee, Jake Carlson, Maribeth Slebodnik, Bert Chapman. "It's in the Syllabus": Identifying Information Literacy and Data Information Literacy Opportunities Using a Grounded Theory Approach[J]. The Journal of Academic Librarianship,2015,41(4).
19. Liu Xun. Exploring embedded data literacy education in higher education libraries [J]. Library and Intelligence,2016(03):86-89.
20. Zhang Chen. Libraries and data literacy education in the ear of big data[J]. Library and Intelligence,2014(04):117-119.
21. Dolezel D, Shanmugam R, Morrison E. Are college students health literate? [J]. Journal of American college health: J of ACH,2020,68(3).
22. Jidkov L, Alexander M, Bark P, et al. Health informatics competencies in postgraduate medical education and training in the UK: a mixed methods study. [J]. BMJ open,2019,9(3).
23. Genes to Society: A Curriculum for the Johns Hopkins University School of Medicine [EB/OL]. [2021-11-01]. https://www.hopkinsmedicine.org/som/curriculum/genes_to_society/index.html.
24. Lei Chunrong, Chen Meng. A comparative study on data literacy education in university libraries at home and abroad [J]. Library,2021(04):47-51.
25. Wang Yuxin, Liu Hongqi, Yu Qi, et al. Survey on data literacy of medical posrgraduates in Shanxi Province[J]. Medicine and Society,2020,33(08):94-97.
26. Wu Aizhi, Wang Sheng. Research on the design of data literacy education system in university libraries-a case of Peking University Library[J]. journal of University Libraries, 2020,38(06):96–103.