



Exploration and Practice of a Training Path for University Students' Data Literacy

Shifen Zhong* and Rongqi Liu
Xihua University, Chendu, China

Abstract

In the age of big data, data literacy has become an essential skill for university students. The effective management of data information is increasingly vital in both scientific research and social decision-making processes. One effective method to fostering data literacy among students is to offer a general education course of big data that is accessible to students from various disciplines. Over four years of teaching the Big Data General Course, we have developed teaching schemes, methods, and evaluation strategies to enhance students' data literacy across the university. Leveraging our experiences at Xihua University, we collected and standardized pertinent student learning data, and organized and visualized the data for analysis. Furthermore, we implemented two types of evaluation methods: summative self-assessment and formative assessment. The results of our analysis and evaluation demonstrated a positive correlation with the intended learning outcomes. Our findings indicate that providing a general course on big data is an effective means of cultivating data literacy among university students. This conclusion provides valuable insights for enhancing general competencies among students in higher education.

Keywords: data literacy, training path, general education, evaluation

1. Introduction

The swift advancement of technologies such as artificial intelligence, cloud computing, mobile internet, and big data continues to amplify societal informatization, tightening the bond between people and data. As far back as 2012, the United Nations published a white paper on big data governance titled "Big Data for Development: Challenges & Opportunities." This aimed to encourage the societal use of big data and stimulate global development, leading to worldwide attention on data management and data literacy education. "Living Knowledge: The British Library 2015-2023" highlights the ongoing revolution in the creation, analysis, and utilization of data in all forms, from vast scientific and social datasets - often labeled as 'big data' - to the innovations derived from analyzing digitized cultural content in the humanities (British Library, 2015). In the era of big data, data types are becoming more varied and complex, making data literacy an increasingly essential quality for citizens.

2. Overview of Data Literacy

2.1. Data Literacy Definition

Data literacy is a subset of scientific literacy and a vital component of information literacy. It involves understanding and applying data, including aspects such as consciousness, ethics, collection, organization, and analysis (Jiao, 2020).

Data literacy comprises three areas: data consciousness, data skills, and data ethics. Data consciousness, the foundation of data literacy, pertains to recognizing data's value and potential harm. Data skills, the core of data literacy, involve the abilities needed for various activities during the data lifecycle. These activities include collection, representation, description, discovery, retrieval, selection, evaluation, analysis, utilization, citation, integration, reuse, preservation, and management (Huang & Li, 2016).

Data ethics, a crucial aspect of data literacy, refers to the moral consciousness, norms, and behaviors regulating social relationships that arise during data utilization. Data literacy also emphasizes technical elements such as data production, preservation, and management (Qin & Chu, 2019). The content of data literacy may vary based on different research perspectives.

In the era of big data, data has evolved into a significant resource in our society. By examining keyword co-occurrence maps and referring to original literature, it is clear that many researchers associate big data with the definition of data literacy. This suggests that big data is the foundation of data literacy's emergence and holds strategic value (Zhao, 2023).

2.2. Importance of Data Literacy in Higher Education

In the era of big data, data literacy has emerged as a crucial skill applicable in our daily lives. G. Jonathan and his colleagues argue that data literacy education can cultivate not only people's sensitivity to data science, but also awareness of data sociology, data politics, and participation in public data infrastructure (Jonathan, et al., 2018).

Firstly, data literacy is vital at all educational levels and throughout lifelong learning (Prado & Marzal, 2013). Secondly, the big data environment's exponential growth makes data literacy a crucial skill for adapting to modern society. Thirdly, by acquiring data literacy, individuals can participate in data production, fostering social knowledge sharing and development. Lastly, good data literacy helps protect personal data privacy and promotes reasonable and legal data usage. Enhancing public data literacy can help reduce or eliminate the data gap between developed and developing countries.

The data literacy level of university students, being the backbone of future society, is particularly important. Thus, in the context of the big data era, fostering data literacy in higher education is indispensable.

2.3. Status of Data Literacy Education

Through a systematic review and inductive analysis of relevant literature, Meng Xiangbao and others (Meng et al., 2016) have categorized the current research on data literacy into four aspects: the connotation of data literacy, the demand for data literacy, data utilization behavior, and data literacy education. Present research on data literacy education reveals some relatively mature systems for implementing such education. These primarily involve the construction of scientific data management resource navigation, general education on data literacy, and discipline-specific data literacy education. Qiuyan Huang and others have described the status of geographic information science data literacy courses in American, British, and Australian

universities. These courses have clear objectives that align with social and disciplinary development, and the course content emphasizes the training of core capabilities in geographic information science (Huang et al., 2020). Luman Wang and others have used the educational practice of "Medical Big Data Analysis" as an example to propose the deep integration of data literacy education within specific disciplines (Wang et al., 2022). The relevant literature indicates that the primary focus of data literacy education is on general data literacy, as well as the integration of data literacy with various disciplines.

3. Data Literacy Training Program for University Students

The objective of data literacy education is to foster citizens' awareness, knowledge, skills, and ethics related to data. We achieve this through a combination of general, professional, and personalized education, focusing on awareness, knowledge, skills, and ethics within the context of the data life cycle.

3.1. The Main Problems of Data Literacy Training

Some studies suggest that data literacy courses should be made a core part of general education in universities to promote the professional development of data literacy education (Zhang & Yang, 2020).

Taking Xihua University as an example, this paper introduces a scheme of data literacy training for university students. Xihua University is a comprehensive university that integrates multiple disciplines such as science, engineering, literature, and management, which provide theoretical and application support for each other. Due to the large differences in students' majors, it is especially important to offer general data literacy courses for all students. The main problems in the process of data literacy training are as follows:

1. After the concept of big data emerged, it became clear that students lacked data literacy. There was minimal overlap between general education and professional teachers' teams, leading to a disconnect between general education content and students' majors.
2. The scope of teaching exchange is confined within the college, leading to a singular cultivation approach. The class content is uniform, the teaching methods are limited, and general education is standardized. As a result, it becomes challenging to tailor teaching according to the students' aptitude.
3. Traditional teaching uniformly uses one textbook, teaching resources are limited, teaching methods are uniform, evaluation methods are inflexible. This results in subpar educational outcomes and fails to genuinely enhance students' abilities.

3.2. Main Methods to Solve the Problem of Data Literacy Training

Xihua University is addressing the issue of limited data literacy among students. Its goal is to foster individuals with skills in data thinking, data abilities, and data ethics. The university is actively exploring and implementing changes in data literacy general education. This is based on a mix of both professional and general education.

Through careful exploration, we've established a data literacy training system. This includes a set of training goals, a tiered general education course group, and a high-quality teaching team. We've promoted general data literacy education integrated with specialized knowledge and innovated a "four combinations" training approach. Additionally, we've built a categorized and layered teaching model tailored to students' abilities.

We've established versatile teaching resources and adopted diverse teaching methods, such as blended and case-led learning. We've also implemented a learning assessment evaluation mechanism focused on the learning process. By integrating theoretical and practical teaching and expanding teaching time and space, we've successfully engaged and piqued students' interest.

3.3. Data Literacy Training Teaching Plan

3.3.1. Formulate Data Literacy Training Goals

In 2019, Xihua University formulated a talent cultivation plan with the fundamental principle of "characterizing general innovation education and continuously optimizing the undergraduate teaching curriculum system." The key areas of revision were "strengthening general education and improving students' scientific and humanities literacy." Data literacy is a crucial aspect of this scientific and cultural literacy. To overcome the limitations of specialized students' data literacy training, we've broadened our goals. They now include cultivating data thinking, data abilities, data security awareness, and ethics, not just data knowledge. For thought enlightenment, we aim to guide students into the big data era. We encourage them to actively engage with big data changes and develop cross-disciplinary thinking abilities. In terms of data abilities, we focus on helping students understand big data concepts and familiarize themselves with its applications. We also aim to explore the intersection of big data and their majors, and inspire a passion for innovation and entrepreneurship based on big data. Regarding big data security and ethical awareness, we emphasize the importance of data security awareness. We encourage students to maintain critical thinking amidst the big data frenzy, understand the safety and ethical issues posed by big data technology, such as privacy leaks, information security, and data divide. We strive to ensure their actions adhere to the ethical norms of big data.

3.3.2. Build a Data Literacy Training System

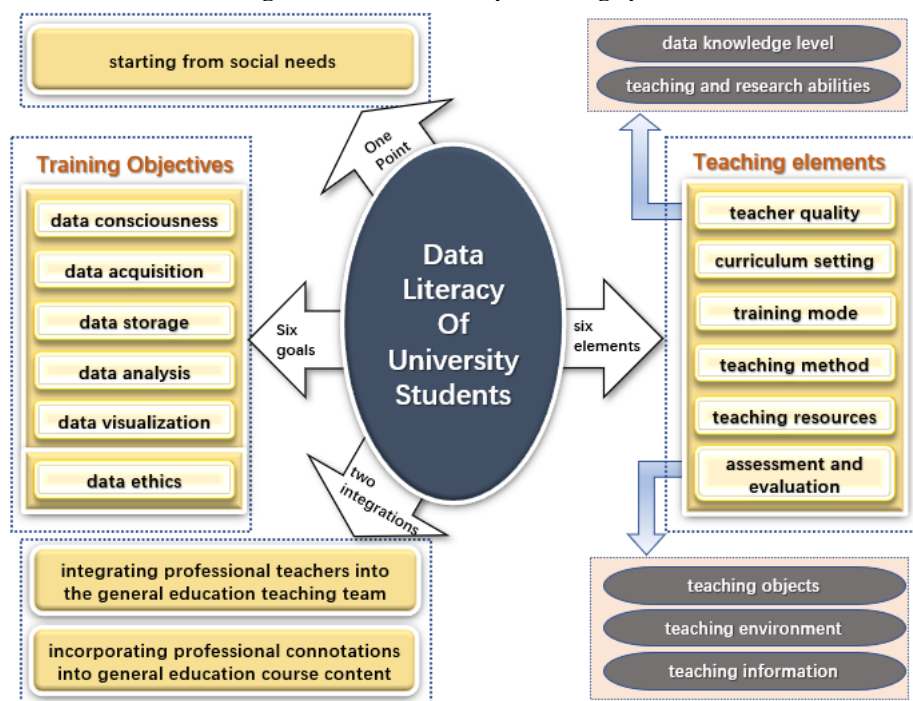
Stay current by updating the data literacy training objectives as needed. Establish a data literacy training system that includes one starting point, two integrations, six objectives, and six elements, as illustrated in Figure 1.

Among them:

The term "one starting point" implies starting from social needs.

The phrase "two integrations" means the inclusion of professional course teachers into the general education teaching team, and incorporating the meaning of different majors into the general education teaching process.

Figure 1: Data literacy training system



Source: Author's own work

The "six goals" are designed to equip students with the cognitive ability to manage small data sets, the awareness and ethical guidelines to handle large data sets, and the innovative ability to analyze big data applications in their specific field from an interdisciplinary perspective. These goals are data awareness, data acquisition, data storage, data analysis, data visualization, and data ethics.

The "six elements" refer to six teaching components: teacher quality, curriculum setting, training mode, teaching method, teaching resources, and assessment and evaluation methods. Teaching resources encompass teaching objects, teaching environment, and teaching information.

3.3.3. Implementing "Specialized-General Integrated" Data Literacy General Education

(1) Collaborate with various academic disciplines to provide data literacy general education courses in a reasonable manner.

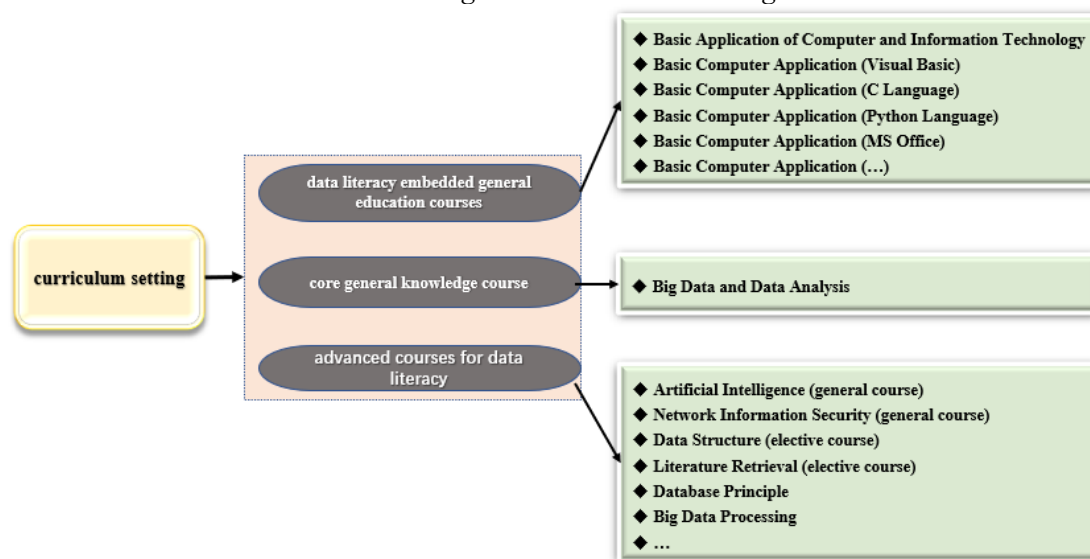
Adjust course settings and content in accordance with societal needs in a timely manner. Gradually, we have developed three categories of courses in stages: data literacy embedded general education course group, data literacy core general education course, and advanced data literacy course group (as depicted in Figure 2).

Data literacy embedded general education courses group, includes the following: Basic Application of Computer and Information Technology, Basic Computer Application (Visual Basic), Basic Computer Application (C Language), Basic Computer Application (Python Language), Basic Computer Application (MS Office: Microsoft Office), and others.

The core general knowledge course for data literacy is Big Data and Data Analysis.

The advanced courses for data literacy include: Artificial Intelligence (general course), Network Information Security (general course), Data Structure (elective course), Literature Retrieval (elective course), Database Principle, and Big Data Processing, and others.

Figure 2: Curriculum setting



Source: Author's own work

(2) Integrate general and specialized knowledge to foster a connection between general education and professional teaching.

A "double integration" strategy is used to tackle the disconnect between general knowledge and professional connotations. This involves integrating professional teachers into the general education teaching team, and incorporating professional connotations into general education course content.

The teaching team's structure is optimized to establish a highly educated general education teaching team. Faculty support is provided by the School of Computer and Software Engineering, Teacher Development Center, and Academic Affairs Office, pooling resources to create a diverse team of professors, associate professors, lecturers, and doctors.

A well-organized teaching team, complete with excellent educators, academic leaders, and individuals excelling in teaching and research, has been established for basic computer education and general courses on big data and data analysis. This team guarantees the incorporation of professional connotations into teaching by advocating for educational reform and competitions, and utilizing research feedback to improve teaching methods.

(3) Innovate the "four combinations" training approach, and establish a classified and tiered training model.

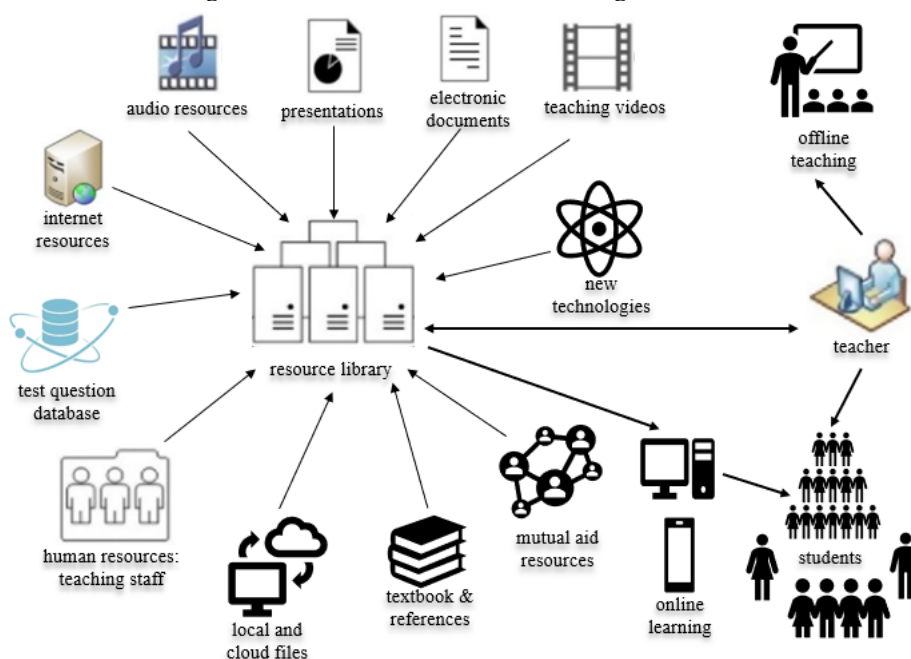
Through the combination of school and school, school and enterprise, college and college, and inside and outside the class, we foster multi-party collaboration and multi-level progress in data literacy general education. Through lectures, competitions, and other extracurricular activities, combined with in-class learning, we enable students to understand data and its applications in professional fields such as legal data, architectural data, traffic data, and financial data. This further achieves an integration of specialized and general knowledge, as well as innovative thinking.

Following many theoretical discussions and practical explorations, we concentrate on one main goal - enhancing students' data literacy. This is categorized into two main areas: science and engineering, and humanities and social sciences. These categories encompass multiple majors, allowing students from all disciplines to understand and use data within their specific field, thereby creating value. The differentiated content on the teaching platform, facilitates layered teaching.

(4) Establish stereoscopic teaching resources, utilize diverse teaching methods, and set up an academic assessment evaluation mechanism that emphasizes the process.

The teaching team has compiled teaching materials and aids like "Big Data General Reading", "Basic Application of Computer Information Technology", "Visual Basic Program Design", and other related course clusters. They also developed the "Big Data and Data Analysis" general education course online teaching platform, "Microsoft Office Advanced Application" MOOC (Massive Open Online Course), "C Programming Language Design" MOOC, "Visual Basic Program Design", "Computer and Information Technology Application Foundation", "Data Structure" school-level boutique course website. The high-quality course platforms, building three-dimensional teaching resources with features of hierarchical teaching objects, digitized teaching environment, and systematized teaching information (as shown in Figure 3).

Figure 3: Three-dimensional teaching resources



Source: Author's own work

3.3.4. Establish a Big Data General Education Course as the Core Curriculum to Enhance Data Literacy of University Students

Today's society has entered the era of big data. Generally, data literacy extends to individuals with comprehensive data skills. These skills are required across various fields, leading to high demand and presenting a significant challenge in cultivating such talents. As shown in Figure 1, we incorporate "Big Data and Data Analysis" as the core general education course for enhancing university students' data literacy in our "specialized integration" data literacy general education process. While this course serves as an introduction for computer science students, its primary target is students from non-computer science disciplines. Establishing big data general education courses is crucial for cultivating big data composite talents.

3.3.5. Teaching Methods for Cultivating Data Literacy

Promote "blended" and "case-led" teaching models when fostering data literacy in university students. Implement a blend of online and offline teaching using multidimensional resources. Connect lessons with real-life scenarios from work, new technologies, and various fields such as culture, marketing, and public health. Extend teaching beyond the classroom and textbooks

through traditional teaching methods, discussions, reports, lectures, and competitions. This approach, applied in the first, second, and third classroom forms, enhances students' data application skills. In the teaching process, emphasize the "case-led" teaching method. This method encourages problem-oriented learning. Guide students to actively explore and solve small and big data-related problems in meaningful situations, leading to concept formation.

4. Student Learning Data Analysis Based on the Core General Education Course on Data Literacy

For the core general education course, "Big Data and Data Analysis", we collected students' basic data and learning trajectories from the learning platform. We standardized these data and conducted a visual analysis to identify key factors influencing the course outcomes. Using the RFM model (Recency Frequency Monetary model), we applied the K-means algorithm to cluster students based on their course progress and established three warning levels. Each level has a corresponding teaching strategy, including personalized teaching plans. Particularly for high-risk students, we devised intervention strategies through data analysis. This approach aims to help more students successfully finish the course and boost their data literacy (Wei et al., 2021).

4.1. Standardize Student Learning Data

In traditional teaching methods, course assessments primarily relied on students' exam scores. This approach overlooked the management of the teaching process and neglected daily performance, homework, and learning trajectories, all of which are crucial to student performance. For the core general education course "Big Data and Data Analysis", we gather related online data from students on the learning platform. This data includes daily attendance, task completion, online learning time, and homework submission. After standardizing this data, the RFM model we built demonstrates that the interval between the student's last online learning session and the homework deadline, the frequency of online learning, and the total learning time are crucial indicators for prediction and classification. These three features significantly impact classification and prediction (Wei et al., 2021).

4.2. Build a Model and Apply it to the Course Warning Case Process

Using the "Big Data and Data Analysis" course at Xihua University as an example, the implementation process mainly includes: data preprocessing, model construction, attribute standardization, cluster analysis, and classification strategy. Data preprocessing involves handling missing values, performing outlier analysis, cleaning, and standardization. Following data preprocessing, a descriptive analysis is performed on the experimental data. Key indicators that may influence students' final scores are selected from the data. This is followed by correlation analysis after normalization and standardization. By using the RFM model, we can promptly identify students' learning difficulties and provide appropriate support and intervention measures, helping them complete the course more effectively (Wei et al., 2021).

4.3. Design Evaluation Method

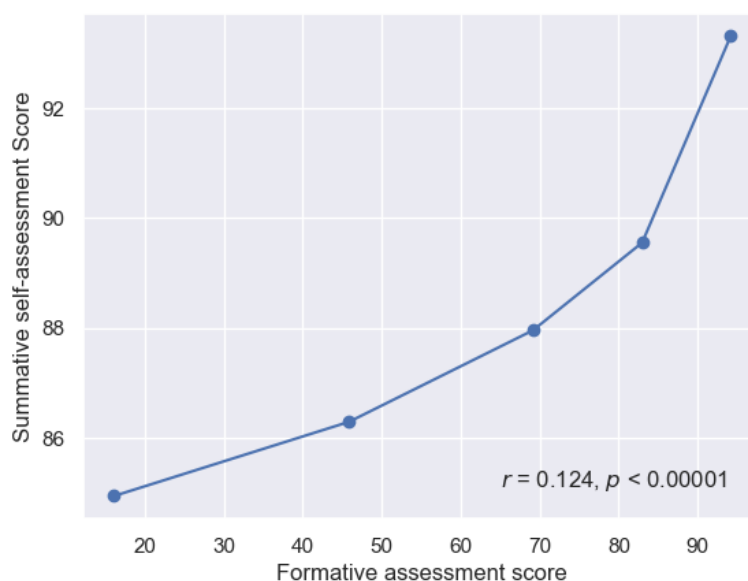
In the core general education course "Big Data and Data Analysis", we explored two assessment methods: summative self-assessment and formative assessment. Our research indicates that summative self-assessment fosters student initiative, a crucial aspect of future evaluation processes. Both these assessment methods have indicators of different importance, which need to be combined to obtain the final score. Upon analysis, while summative self-

assessment aligns with our objectives for the course, it does not fit the students' grading standards. Formative assessment, being objective, can serve as a grading standard (Fu et al., 2021).

4.4. Evaluation Results Analysis

We utilize the entropy weight method along with various indicators for comprehensive self-assessment and formative assessment. This yields the final formative assessment score or comprehensive self-assessment score for each student. Subsequently, we calculate the Pearson correlation coefficient of these two scores. As shown in Figure 4, the correlation coefficient indicates a positive and statistically significant correlation between the two scores (Fu et al., 2021). Summative self-assessment, which involves personality and its connection to the profession, aligns with our teaching objectives. However, it is not suitable as a grading standard due to the potential for student dishonesty. In contrast, formative assessment, composed of objective indicators, can be used for grading. In practice, both methods can be employed together. Self-assessment checks if teaching objectives are met, while formative assessment scores contribute to the student's course grades. As a result, the course grades for our core general education course "Big Data and Data Analysis" in data literacy consist of formative assessment scores and final grades (Fu et al., 2021).

Figure 4: Relation between objective scores and subjective scores



Source: (Fu et al., 2021)

5. Conclusion

Big data, a pivotal force across multiple facets of society, is rapidly expanding and evolving, emerging as a strategic asset for the future. In the age of big data, the capacity to swiftly acquire, analyze, and harness diverse datasets is paramount for personal advancement. Thus, it is imperative for individuals to bolster their data literacy by augmenting their awareness, knowledge, and skills.

University students, poised as the future catalysts of societal progress, stand to benefit significantly from robust data literacy. To nurture this competence among students, integrating interdisciplinary data content into general education curricula and leveraging a blend of online

and offline pedagogical approaches can foster a holistic grasp and utilization of data across academic disciplines.

Systematically collecting and analyzing data on student learning experiences not only heightens student engagement but also furnishes an objective assessment of their educational achievements.

Considering the diverse array of university disciplines, our research underscores the efficacy of instituting a suite of data literacy training courses tailored to individual majors, alongside offering overarching data literacy programs. This strategic approach serves as a blueprint for enhancing the comprehensive skill sets of university students.

Acknowledgment

This paper is an output of the Xihua University's school-level first-class course construction project "Big Data and Data Analysis", the Xihua University's school-level course ideological construction project "Big Data and Data Analysis", the Sichuan Province education and teaching reform projects (No: JG2021-929), the Xihua University education and teaching reform projects (No: xjjg2021115 and xjjg2021049), and this work supported by China Scholarship.

References

- British Library (2015). Living knowledge:the British library 2015-2023. [Online]. pp. 6. Available: <https://www.scribd.com/document/252532741/Living-Knowledge-the-British-Library-2015-2023>
- Fu, X., Zhong, S., He, Z., Tang, J., Chen, H. and Wei, D. (2021). Research on Summative Self-assessment and Formative Assessment for Big Data General Course. *2021 IEEE 3rd International Conference on Computer Science and Educational Informatization (CSEI)*. Wulumuqi, China, pp. 254-258.
- Huang, Q., Liu, J. and Deng, S. (2020). "Characteristics and its enlightenment of data literacy of GIScience postgraduate curriculum in universities among USA, UK and ASU", *Higher Education Forum*, vol. 4, pp.21-27.
- Huang, R. and Li, B. (2016). Data literacy education: expansion of information literacy instruction in the big data era, *Documentation, Information & Knowledge*, vol. 1, pp. 21-29.
- Jiao, J. (2020). Research on the data literacy education in university libraries in the age of big data, *Library Work and Study*, vol. 3, pp. 81-84.
- Jonathan, G., Carolin, G. and Liliana, B. (2018). Data infrastructure literacy, *Big Data & Society*, vol. 5, pp. 1-13.
- Meng, X., Chang, E. and Ye, L. (2016). Data literacy research: origins, progress and prospects, *Journal of Library Science in China*, vol. 2, pp. 109-126.
- Prado, J.C. and Marzal, M.A. (2013). "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents", *Libri*, vol. 63, pp. 123-134.
- Qin, X. and Chu, J. (2019). Research on the connotation structure of scientific data literacy, *Library and Information service*, vol. 63, pp. 30-39.
- Wang, L., Qi, H., Wang, J. and Wang, C. (2022). "Research on Data Literacy Education for Special Fields: Taking the General Course Construction and Teaching Practice of Medical Students as an Example", *Library*, vol. 7, pp. 15-21.

- Wei, D., Zhong, S., He, Z. and Tang, J. (2021). Construction and Practice of Curriculum Pre-Warning Model based on Data Analysis. *The 16th International Conference on Computer Science & Education (ICCSE 2021)*. Lancaster, United Kingdom, pp. 286-291.
- Zhang, J. and Yang, W. (2020). Research on Foreign Data Literacy Education and Its Enlightenment, *Digital Library Forum*, vol. 4, pp. 60-67.
- Zhao, F. (2023). Research scheme and prospect of domestic data literacy education, *Library Theory and Practice*, vol. 4, pp. 30-40.