Research on Scientific Data Literacy Education System

Zhou Qi

Library of Jinan University, Guangzhou, China
Email: tibcf_zhou@jnu.edu.cn

Abstract

This article analyzes the connotation of scientific data literacy, uses literature research methods and case analysis methods to investigate and analyze the scientific data literacy education practices of universities or institutions, and proposes the scientific data literacy teaching objectives, teaching objects, education models and education strategies. It has important theoretical value and practical reference.

Keywords
Science Data Literacy, Teaching Objective, Embedded Education, MOOC, Education Model, Education Strategy, Teaching Evaluation

1. Introduction

Under the E-Science environment, the paradigm of scientific research has changed. Following experimental science, theoretical science, and computational science, a fourth research paradigm called “data-intensive science” emerged. Under this type of scientific research model, “data drives scientific development, science is data, and data is science” [1]. Scientific data is not only the product of scientific research activities, but also the foundation of scientific research activities. It contains a lot of research value and use value. Scientific data has become an important academic information resource. Researchers have explored new scientific research results through the analysis of a large amount of scientific data and promoted scientific development.

The formation of an intensive data environment has caused the scientific data produced by universities to grow in volume in terms of quantity, variety, and speed. Researchers face a series of data management issues, such as data management planning, data citation, data publishing, and ethical use of data, etc.
Scientific data literacy, as a key concept of data management, has become one of the necessary capabilities for academic researchers to research and communicate. US National Natural Science Foundation, National Institutes of Health and other research funding agencies, and data management and sharing policies of funding agencies such as the UK Higher Education Funding Council, the UK Research Council, Welcome Funds, and Research Information Network, also put forward requirements for researchers’ data management capabilities. In 2010, the 76th IFLA Congress was held in Gothenburg, Sweden, and social science data literacy became one of the conference themes \[2\]. In 2012, the Institute of Museum and Library Services (IMLS) funded the “Data Information Literacy Project” to explore how to cultivate the ability of the next generation of scientists to find, organize, utilize, and share data \[3\].

It can be seen that the problem of scientific data literacy has aroused the attention of the library industry. For university libraries, cultivating the scientific data literacy of teachers and students is an urgent and important task as well as a new opportunity and challenge. Some academic libraries in Europe and the United States have already carried out appropriate literacy education activities to increase scientific personnel’s data awareness, data collection and analysis skills, and promote scientific data management and sharing.

2. Research Review

From the perspective of literature research, the foreign library community has focused on the role and methods of library scientific data literacy training. For example, Koltay emphasized the importance of data literacy in the mission of university libraries, and pointed out that data literacy should have a unified term \[4\]; Frank et al. introduced the meteorological disciplines as an example to introduce the data literacy education of university libraries \[5\]; Dillo paid attention to the joint data management of archives and libraries. Training activities, the establishment of the FrontOffice-BackOffice model \[6\]; Maybee using a grounded theory to review the syllabus of the nutrition science and political science of Purdue University, analysis of students’ information literacy and data information literacy demand \[7\]. Most foreign research libraries have already carried out corresponding scientific data literacy education activities and diversified their teaching forms, including offering elective courses, seminars, and online courses, such as Harvard University, Yale University, and University of Virginia. The University of Virginia Library’s data literacy education curriculum is well-established. It is designed longitudinally according to the data life cycle, and is rolled out horizontally in different subject areas to provide specialized training for researchers in specific disciplines.

Chinese university libraries have realized the importance of scientific data management and have begun exploring data management in the construction of institutional knowledge bases such as Peking University, Tsinghua University, Xiamen University, Xi’an Jiaotong University, and Wuhan University. In 2014, led by Fudan University, nine domestic university libraries jointly initiated the
establishment of the “China University Library Research Data Management Promotion Working Group”. In December of the same year, the social science data platform of Fudan University was formally launched to provide universities, research institutes, and government agencies with functions for storing, publishing, exchanging, sharing, and online analysis of scientific research data. Some university libraries in China try to implement the scientific data literacy education practice, which is mainly reflected in the training of data statistics databases and analysis software.

3. Scientific Data Literacy Education System

“Scientific data literacy is the knowledge and ability to collect, process, manage, evaluate, and use data in scientific research. Although scientific data literacy is similar to information literacy and digital literacy, it focuses on data collection, processing, management, evaluation, and use. The multiple capabilities, rather than literature, emphasize the ability to generate, manipulate, and use datasets in scientific research” [8]. Jacob Carlson [9] believes that the basic content of information literacy education includes the introduction of database and data formats, data discovery and acquisition, data management and organization, data conversion and interoperability, data quality assurance, metadata, data management and reuse, data preservation, data analysis, data visualization, data ethics (including data references), etc. The core competencies of scientific data literacy include understanding data, interpreting and evaluating data, managing data, and using data [10]. From the above analysis, scientific data literacy is similar to information literacy, including data awareness, data management knowledge, and data management skills. At the same time, scientific data literacy has a periodic nature, emphasizing the collection, processing, and evaluation of scientific data. Management and utilization activities focus on the various skills required to manage data that are required in the basic scientific research process. In addition, scientific data literacy emphasizes the ability to analyze data, present data, and use data management tools.

Specific to the subject area, the requirements for scientific data literacy ability are more specific and more specific. For example, sociology emphasizes the ability of data collection and statistical analysis, and economics specializes in designing quantitative economics courses, emphasizing data analysis and modeling capabilities. Bioinformatics emphasizes the ability to use computer as a tool to store, retrieve, and analyze biological information. In the field of journalism, the Digital News Center of the School of Journalism at Columbia University has targeted the new position of “data journalist” and proposed that the six hard skills that journalists in the post-industrial era should possess include data and statistical capabilities, and master user analysis, tool capabilities and data analysis skills [11]. The data literacy in the subject area has embedded characteristics. This embeddedness is reflected in the cooperation of teaching methods. For example, teachers from the Department of Sociology at the University of California, Los Angeles, collaborate with librarians in data lite-
racy education [12], and professional teachers teach scientific research methodologies with professional knowledge, the librarians impart skills such as data collection, storage and management, and work together to give full play to their respective advantages.

3.1. Scientific Data Literacy Teaching Objective

Researchers face new requirements for managing, sharing and supervising data, so having the appropriate knowledge and skills is crucial. However, data management ability education is often not used as the content of the student’s curriculum system. The disconnection between scientific data management needs and researcher data management capabilities has become a major obstacle to data protection and the full use of scientific data. Most of the university library’s data literacy education has clear goals. For example, the goal of Purdue University’s data literacy program is to establish the foundation of library science data literacy education, and to develop appropriate librarians based on the appropriate data management skills in the subject area. Data literacy courses and programs provide standard processes. The MIT Library’s data management training goal is to help researchers learn scientific data management skills, including the basics of scientific data management, scientific data file organization, and version control. The UK Data Management Center (DCC) believes that under the e-Science research environment, research libraries conduct data management training and data literacy education to help researchers arm their knowledge in data sharing, preservation, and long-term access skills, and as a goal to guide the development of various activities.

3.2. Scientific Data Literacy Teaching Object

Graduate students and researchers are the main objects of data literacy education in university libraries. Graduate students are the bearers and successors of future scientific research work. The problem they face is how to change from a student role to a professional field researcher. It is of great significance for graduate students to adapt to study life and participate in scientific research practice. It is also indispensable for researchers who have engaged in front-line research to receive data management training. They are faced with huge data processing and preservation issues. How to develop data management plans to meet the data management and sharing requirements of funding agencies and publishers, how to manage data in a standardized manner for future discovery and reuse, how to implement data security and backup and long-term preservation, and how to comply with data ethics and ethics. Problems require the librarians to provide the necessary assistance and support.

3.3. Scientific Data Literacy Teaching Content

Scientific data awareness education. Although scientific data management is being familiarized and understood by more and more researchers, data management is in many cases operating at the institutional level rather than the needs of
the users of the academic community, in addition to the data storage, the demand for user services. The response is actually limited or missing. Therefore, for researchers, data literacy education should first be ideological education, that is, to allow researchers to understand the basic terms, key concepts, policies, roles, personal roles and responsibilities of data, and to cultivate user needs.

Scientific data literacy skills education. Teach data resources and data analysis software for different subject areas, provide data management guidance around the life cycle of scientific research projects, make researchers familiar with the data life cycle, and have the knowledge and skills to find, analyze, manage, utilize and share data in the subject areas.

Scientific data literacy application education. Emphasis on the ability of researchers to data mining, data discovery, that can use data resources to find problems, analyze problems and solve problems, and even create data management plans for specific research projects, which focuses on skills and practice training.

Scientific data ethics education. To develop the ability of scientific researchers to use data scientifically, including data citation, intellectual property and copyright, privacy and security issues. In the current situation where the proliferation of academic utilitarianism and the lack of academic integrity occur, the cultivation of noble data ethics is particularly important.

Innovation ability improves education. It is mainly to help researchers master the methods of scientific data quality screening, evaluation, clustering, correlation, integration, rediscovery, etc., to identify valid, novel, potentially useful data from the data set and integrate it into their own existing knowledge so as to realize Knowledge innovation, research innovation.

3.4. Scientific Data Literacy Education Model

1) Embedded in scientific research activities for data literacy education

The embedded scientific data literacy education model embeds the content of scientific data literacy education in the teaching of specialized courses and regards the library scientific data literacy education as an integral part of the curriculum objectives of various subjects. It not only completes professional teaching, but also requires students to master the knowledge of scientific data management. And skills, and use scientific data to solve professional problems. The University of California, Los Angeles Library embeds data literacy and information literacy education in a sociology program that is coordinated by a librarian and a data archivist based on syllabus [13]. The DMTpsych Data Literacy course at York University helps psychology researchers and researchers learn data management and develop data management plans [14]. The Purdue University Library has conducted research on data literacy education for researchers in fields such as agriculture and bioengineering. The members are composed of data librarians, subject specialist librarians, and members of the discipline [15]. The Data Train of the Cambridge University Library Data Literacy Education Project is targeted at archeologists and socio-anthropologists [16]. It can be seen
that the embedded scientific data literacy education model requires that library data librarians cooperate with professional teachers, jointly conduct curriculum and lesson plan design, and undertake scientific data literacy education module teaching. This puts high demands on the professionalism and understanding structure of library data librarians.

2) Whole-course education based on the life cycle of scientific data

The scientific data life cycle stems from the life cycle of scientific research and can be broadly divided into five stages: data acquisition, data production, data storage and management, data preservation and sharing, data citation and publication. The library can develop a full-process data literacy education model based on the life cycle of scientific data. The “data collection phase” is the start-up phase of scientific research projects. This phase is mainly to provide training and lectures on data resources and provide data support for project initiation. This is the strength of university libraries. At this stage, the library can introduce the basics of data management and assist in the development of data management plans. The “data production phase” and the “data storage and management phase” run through the entire research phase of the scientific research project. This phase produces a large amount of raw data and experimental or survey data, and stores and manages the data. The library can be integrated into research. In the team, explain data statistics, how to use analysis software, conduct data analysis and use metadata to describe the professor of data sets. At the same time, data management is performed according to the original data management plan. This requires higher academic background and scientific research ability for librarians. It also requires a large amount of investment by university libraries to obtain sufficient hardware and software support, including technical support, hardware equipment, and personnel strength. The “data preservation and sharing” stage is to fully protect and utilize the scientific data produced. University libraries have unique advantages in this regard and should assume their own responsibilities. The “data reference and publication phase” is the final stage of the scientific data life cycle. At this stage, university libraries mainly carry out education on data reference academic norms, which can increase data reusability and sharing ability, and increase the recognition of data producers. The verification of the scientific research process is the proper meaning of the university library. The university library can explain the data ethics and citation normative knowledge through specific cases, and can also train data mining and data correlation technologies to promote the reuse of data. Make the most of the data. The education model is suitable for “minority” education, that is, a specific project group to participate in a specific project. The education model is also a kind of “practical” education, focusing on the ability to use data resources to solve practical research problems in the scientific research process.

3) Use MOOC mode

In recent years, the rise and development of the mass open online course (MOOC) has brought about a positive and significant impact on the global
higher education and has become an effective complement to higher education. MOOC is learner-centered and provides learners with personalized education services that fully stimulate learners’ subjective initiative and improve their learning effectiveness.

Facing the reform of educational technology brought about by the rise of MOOCs, and the reform of educational concepts and teaching models, we should make full use of the current network environment and information exchange technologies, effectively take advantage of the MOOC platform, and seize the opportunity of providing embedded education to break professional teaching. The boundary between scientific data literacy education and the professional teachers’ development of concept maps, design team projects, discussions, and assignments, and the integration of scientific data literacy-related teaching contents at any time in accordance with the professional curriculum system, providing learners with professional-related learning in due course, resources, information methods and other aspects of help and support to achieve professional teaching and scientific data literacy mutual penetration and improvement. This will enable learners to gradually improve their information utilization and research innovation capabilities while mastering professional curriculum knowledge.

3.5. Scientific Data Literacy Education Strategy

1) Develop related data literacy education according to disciplines

Discipline is the most natural and basic basis for differentiating user groups in university libraries. Researchers of different specialities have different needs for scientific data literacy skills. For example, engineering disciplines are based on experimental data and can focus on experimental data statistics, analysis, and visualization of data. Research on humanities and social sciences mainly relies on survey data, government-disclosed statistical data, etc., and can focus on the collection of data and education on evaluation capabilities. Specialization features are especially reflected in the embedded education model based on interdisciplinary cooperation. If you do not have professionalism and do not design tutorial content according to professional characteristics, the embedded education model will lose its significance and effect. Therefore, the library’s scientific data literacy education should emphasize the characteristics of disciplines, differentiate the design of scientific data literacy education curriculum content according to the different needs of the discipline users, and develop data literacy education for researchers in specific disciplines.

2) Conduct heuristic education to develop a comprehensive understanding of the data

The cultivation of data literacy should be based on rigorous scientific research purposes, using the data life cycle as a guideline, heuristic education as a means, and data manipulation to foster a comprehensive understanding of the data. The capabilities covered by data literacy are closely linked to the data life cycle. The data management training education generally adopted by foreign universities is
also based on the data life cycle. Athanases pointed out that the data literacy education and data services provided by university libraries should use the complete data life cycle as the guideline to cultivate users’ understanding of the data life cycle in a series of processes from data acquisition to conversion application, and in the course of data management. Deepen understanding [17]. In addition, the education based on data life cycle should also focus on cultivating critical thinking of users. The reason is that the data collection and transformation are carried out under certain data processing conditions. The data is limited because of the limitations of data processing conditions. The ground will produce different degrees of error, and the critical thinking of the data can also be simply understood as the data provider must be responsible for the data he provides, so that the data provider must timely review the data life cycle, under the appropriate conditions for data Re-acquisition, correction or deletion to ensure the rationality of relevant data research. In terms of specific educational methods, critical thinking also to some extent reflected in the understanding of data sources, including how to properly produce, read graphics and charts, draw correct conclusions from existing data, identify data is misunderstood and Improper use and so on. Under the circumstances that the scientific research environment is constantly changing, university libraries must fully understand the urgent needs of scientific researchers in data management, combine data life education with data literacy education, and use heuristic education to help researchers manage and use data in order to cultivate their good data literacy.

3) **Provide personalized data literacy education courses**

According to the different needs of different users to provide corresponding scientific data literacy education, including:

Scientific data education model for most users. The information quality education course for undergraduates is the main field of scientific data general education. It mainly introduces the basic theory and methods of scientific data, enabling learners to understand scientific data and gradually cultivate data awareness.

Discipline data literacy education model for high-level users (teachers, doctors, masters, etc.). It mainly provides specialized lectures and trainings for specific disciplines. It can rely heavily on subject librarians and rely on academic services and academic platforms. The massive growth and the urgent need to develop scientific data have also added new content to the discipline services, and they complement each other. For example, the Wuhan University Library has embedded a scientific data management module in its academic service platform, which has played an exemplary role for Chinese university libraries.

Personalized scientific data education model for specific users. The development of big data enables scientific data education to be truly personalized. The library can build an online scientific data literacy education platform based on adaptive learning technology, realize user self-organization learning, self-adjustment learning, and can push learning based on user characteristics. Resources, and follow up in time, automatically adjust. Reference consulting services are also an
important means for achieving personalized scientific data education.

Scientific data integrity education model for scientific researchers. Integrate scientific data literacy education in research integrity education, focusing on the importance of scientific data for scientific research for scientific researchers, avoiding academic misconduct caused by data fraud, enhancing the reliability of scientific research results, and opening up new areas of academic norms. Improve the construction of scientific research integrity system. It can thus be seen that the scientific data literacy education in the library is not isolated, but rather chimeric, and the library’s other services work collaboratively.

4) Focus on cultivating researchers’ scientific research capabilities

Scientific research ability is an indispensable and important part of scientific data literacy education. Therefore, scientific data literacy education should focus on cultivating the ability of researchers to conduct scientific research. Knowledge can be taught, and the ability must be obtained by the student through hands-on practice, exploration and summary under the guidance of the teacher. After studying basic theoretical knowledge related to scientific data literacy education, researchers must participate in scientific research projects to cultivate data collection, statistical analysis, management, and data security capabilities in scientific research. The whole-process education model based on the life cycle of scientific data is designed specifically on the basis of this feature. Through the whole process of scientific data life cycle education, researchers can firmly establish the awareness that data management is a necessary part of the scientific research process, in order to realize breakthroughs from data knowledge to data discovery and data innovation, and from students to real students. The transformation of senior professional scientific research personnel can realize the leap from engaging in scientific research under the guidance of mentors to creating new research fields independently.

3.6. Teaching Evaluation

The evaluation of data literacy courses mainly includes two aspects: teacher teaching and student learning outcome evaluation. The evaluation of teachers’ teaching includes evaluation of the methods and models used in the teaching of data literacy to determine whether the expected goals of data literacy education have been achieved. For example, the Purdue University Library conducts focus group discussions after the end of course to assess the course results. The assessment of student learning outcomes is the evaluation of the improvement in data management knowledge and skills through data literacy teaching. For example, Cornell University and Cambridge University Library use the end-of-term assessment, learning achievement report (usually papers and assignments) to inspect students learning effect.

4. Conclusions/Suggestions

Based on the comprehensive research and practice, we can see that scientific data literacy education has risen in foreign research institutions, and domestic re-
search has started relatively late. It is still in the stage of introduction, absorption, and digestion. In general, the following problems exist in the scientific data quality education of university libraries:

1) The theoretical level of the concept of scientific data literacy in libraries is still in its infancy, and no specific theoretical framework has been formed, and the design of specific paths is lacking.

2) Insufficient understanding of the importance of scientific data literacy education, lack of top-level design and policy support.

3) Some libraries have conducted scientific data education practices, but they lack extensive research on the actual needs of users, lack research on the “inter-institutional” collaborative development management mechanism, and lack specialized organizations and professional staff.

4) The library community lacks standardized consensus and collaborative management of scientific data management, scientific data literacy education, etc. There are no uniform standard and corresponding rules in the industry. Each library often sets its own rules and self-contained system, which is difficult to carry out in-depth and effective external cooperation and exchange.

With regard to the problems existing in the practice of scientific data literacy education in university libraries, the scientific data literacy education in university libraries is urgently needed and we can carry out the following tasks:

1) Designing scientific data literacy education courses based on user needs

   The first is to conduct a literature survey on the theoretical and practical status of scientific data literacy education in university libraries, sum up experiences, analyze problems, and explore the feasibility and teaching model of scientific data literacy education in university libraries. Second, through the design of questionnaires, or semi-structured interviews to understand the actual needs of users. Only on the basis of full investigation and study, the basic theories of scientific data literacy education in university libraries can be scientifically and reasonably designed to meet the needs of users in the scientific data literacy education courses in university libraries, and then to meet the needs of users to achieve services conversion of user concept.

2) Multi-party cooperation to build an “inter-institutional” collaborative development management mechanism

   The scientific data literacy education in colleges and universities should proceed from the internal teaching and research work flow of the university, take the library as the leading factor, and take the management mechanism of multi-cooperation and “cross-organization” collaborative development. In this regard, many libraries provide many successful experiences and useful references. For example, Cornell University’s Research Data Management Group (RDMSG) is a multi-academic collaborative effort to help create and implement data management plans, apply best practices to manage data, and provide data management services at any stage of the research process. The Oxford Institute for Emerging Institutional Data Monitoring Services (EIDCSR) consists of several research institutes within the University of Oxford including the University
Computing Center (project hosting, research and consulting), Research Service Office (policy research), Podlin Library (metadata management) and scientific research project team (participating in research) work together to complete. Therefore, in the construction of a scientific data literacy education platform, the library shall collaborate with the school information construction center, and actively cooperate with the relevant departments of the school to improve the functional modules of the platform and achieve as convenient and interactive as possible. In addition, in the design of data literacy teaching programs, the library can work with professional teachers throughout the process to embed data literacy education content into professional curriculum design and teaching practices, share teaching tasks, and conduct education on subject embedded science data.

3) Organization and staff guarantee

The construction of the staff team and organization is the precondition for the library to carry out scientific data literacy education. Many university libraries have set up specialized data management agencies and set up relevant positions based on their responsibilities and service priorities. The New York University Library has established a data service studio and has set up 5 positions: the data service coordinator, Data Services and Public Policy Librarians, Data Services Assistant Librarians, Data Services High Commissioner, and Data Services Librarians. The Johns Hopkins University Library’s Digital Research and Supervision Center has seven positions: Senior Data Education Expert, Academic Communication Expert, and so on. The construction of personnel and organizations for scientific data education should be strengthened. On the one hand, the training and re-education of data service librarians should be strengthened on the basis of existing subject librarians and consulting librarians to form a powerful specialized data education hall, staff team. On the other hand, we must strengthen the construction of library professional data service organizations, coordinate and coordinate the library’s data services, and under the guidance and guidance of the collaborative development management mechanism, gather the strength of all parties to expand library data management and service.

Scientific data literacy education is a new field for university libraries, not only across disciplines but also across traditional university library organizations. University libraries should adapt to the development of the data era, actively change their concepts, fully recognize the importance of scientific data, attach importance to the development of data librarians, and strive to explore scientific data literacy education models and educational content, and gradually establish awareness of data cultivation as a guide. Based on data knowledge and skills, a more complete innovation system of scientific data literacy education for the purpose of scientific data standard application is demonstrated, highlighting the new mission and new value of university libraries in the E-Science environment.
Project

Guangdong Education’s Characteristic Innovation Project (2014GXJK009); Guangdong Education Youth Innovative Talents Project (Humanities and Social Sciences) (2014WQNCX010).

References


https://doi.org/10.2218/ijdc.v8i1.254

http://archaeologydataservice.ac.uk/learning/DataTrain