International Core Journal of Engineering Volume 5 Issue 9, 2019
ISSN: 2414-1895 DOI: 10.6919/ICJE.201908_5(9).0011

Research on the Necessity of Data Literacy Cultivation Based on Data Lifecycle

Xiaorong Hou

College of Medical Informatics, Chongqing Medical University, Chongqing 400016, China

xiaoronghou@cqmu.edu.cn

Abstract

This paper discusses the necessity and importance of cultivating data literacy in data lifecycle. It sorts out the current situation of data literacy training. At the same time, the paper points out that in different disciplines, it is not necessary to emphasize the unified standard of data literacy training, but should explore the core data literacy according to the discipline attributes, so as to apply it to the data lifecycle of scientific research.

Keywords

Data Literacy, Data Lifecycle, Framework, College Students.

1. Introduction

The United Nations officially used the term "data literacy" in its documents issued in December 2014, and ACRL, the most influential institution in the field of information literacy education, has repeatedly used this term. "Data literacy" refers to the ability to acquire, analyze, process, utilize and display data effectively and appropriately with data awareness and data sensitivity, and to think critically about data. It is an extension of statistical literacy and information literacy, a further improvement and deepening of traditional "information literacy" education, and has become a necessary skill for scientific researchers.

The process of data collection (or formation), processing, preservation, dissemination, retrieval, access and utilization to disappearance or no longer being used is the lifecycle of data. In the digital scientific research environment, "lifecycle" is a cyclic process, in which digital resources can be found and reused. The lifecycle of scientific research data generally includes eight stages: data management plan, data generation and collection, data management and organization, data processing and analysis, data storage, data publication and sharing, data discovery and acquisition, data reuse.

Most of the world's research institutes integrated data literacy education into specific subject culture, and embedded in laboratory practice. Based on the data lifecycle, researchers can realize that data management is an indispensable formal link in the scientific research process.

2. Literature Review

In 2007, the National Science Foundation (NSF, https://www.nsf.gov/) funded a data management course designed by Professor Qin Jian of the Institute of Information Studies of Syracuse University to improve students' data literacy in the e-Science environment[1]. Since 2011, NSF required that all applications for funding projects must be accompanied by a standardized data management plan to achieve scientific storage, preservation and management planning of scientific data types, formats, records, documents, metadata, etc[2].

In 2012, the Institute of Museum and Library Services (IMLS) funded the "Data Information Literacy" project, which aims to develop the ability of future researchers to find, organize, process and share data, and considered the information literacy commonly referred to as "21st Century

ISSN: 2414-1895 DOI: 10.6919/ICJE.201908_5(9).0011

Skills"[3]. In the same year, President Barack Obama launched a series of educational innovation projects and plans, starting with college education, to organize training in data knowledge and processing skills, the core content of which is data literacy education.

In 2013, Harvard University established the Research Data Collaborative (RDC) project to provide data literacy education for Harvard University researchers, which has been renamed as the Research Data Management Program. This program connects members of the Harvard community to services and resources that span the research data lifecycle[4]. Meanwhile, it also constructed a tool called Dataverse, which is Harvard's open online repository for sharing, preserving, citing, exploring, and analyzing research data[5].

In 2014, Nature launched the Open Access Journal Scientific Data, which mainly publishes scientific data, and Elsevier created the Genomics Data. The requirement of scholar's data literacy is higher and higher for publishing results in data journals.

The Framework for Information Literacy for Higher Education[6], formally concluded by Association of College and Research Libraries (ACRL) in 2015, can be regarded as one of the most perfect data literacy education standards at present. The framework takes "Authority Is Constructed and Contextual", "Information Creation as a Process", "Information Has Value", "Research as Inquiry", "Scholarship as Conversation" and "Searching as Strategic Exploration" as its key content, further clarifies the basic objectives and direction of data literacy system construction in American higher education. For example, the School of Information Science at the University of Illinois at Urbana & Champaign (UIUC) has set up a course system of Data Curation Education Program (DCEP) [7]. Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities across the sciences, social sciences, and the humanities. Data curation activities enable data discovery and retrieval, maintain data quality, add value, and provide for re-use over time. This new field includes representation, archiving, authentication, management, preservation, retrieval, and use.

3. Training Framework

The uniqueness of data literacy training around project lifecycle lies in: the cultivation of this quality does not exist independently, but depends on the start, operation and end of the project. Therefore, the best form of cultivation is project-driven, that is, the mastery of specific ability of data literacy is fully embedded in the process of actual projects. At this time, the requirements of the project-level should be quasi-diversified, advocating project entry at different levels, and the difficulty of cultivation should be from shallow to deep.

It can be found that every process of the data lifecycle can correspond to data literacy. Therefore, correctly understanding the lifecycle of scientific research data, grasping the data characteristics of each stage of the lifecycle and cultivating the data literacy ability are the prerequisites and foundation for the standardized management of scientific research data and the maximization of data value. Meanwhile, data lifecycle is inseparable from scientific research workflow, which is embedded in every link of project initiation, project implementation and project conclusion. Therefore, to some extent, the data lifecycle can be represented as the project lifecycle.

If we compare the project lifecycle, data literacy definition and data literacy abilities, we may make the following training arrangements. Therefore, a data literacy training framework for college students based on lifecycle theory is being constructed. In different stages of project lifecycle, it corresponds to different data literacy concepts and specific data literacy capabilities. See table 1.

Table 1 Training Arrangement

Project Lifecycle	Data Literacy Definition	Data Literacy Abilities
Launching	Data Awareness	Data Discovering

ISSN: 2414-1895 DOI: 10.6919/ICJE.201908_5(9).0011

		Data Acquisition
Functioning	Data Abilities	Data Organization
		Data Processing
		Data Analysis
		Data Evaluation
		Data Storage
Dormancy Stage	Data Ethics	Data Security
Recovery		Data Sharing
		Data Reuse

However, Table 1 only discusses the core definitions and specific capabilities of data literacy in the project lifecycle, and does not discuss how to develop data literacy around the project lifecycle. Many universities, research institutes and even companies have developed and opened many courses and learning software to improve data literacy, using virtual laboratories, learning groups and other forms to develop data literacy.

We take the National Network of Libraries of Medicine (NNLM) as an example. NNLM in the United States has opened up special data literacy course materials, data literacy guides and reference resources[8]. Activities are specific outreach sessions that may include site visits or training or demonstration of NLM resources. The goal of the NNLM is to advance the progress of medicine and improve public health by providing U.S. health professionals with equal access to biomedical information and improving individuals' access to information to enable them to make informed decisions about their health. Its Biomedical & Health Research Data Management for Librarians is an online training course (https://nnlm.gov/class/biomedical-health-research-data-management-librarians/14877). This course provides basic knowledge and skills for librarians interested in helping patrons manage their research data. Attending this course will improve ability to initiate or extend research data management services. see Table 2.

Table 2 Course Schedule

Objectives	To provide an introduction to data issues and policies in support of developing and implementing or enhancing RDM training and services. This material is essential for decision-making and implementation of these programs, particularly instructional and reference services.	
Topics	an overview of data management, choosing appropriate metadata descriptors or taxonomies for a dataset, addressing privacy and security issues with data, and creating data management plans	
Module 1	Research Data Management Overview & Data Lifecycle	
Module 2	Data Curation & Documentation	
Module 3	Data Standards, Taxonomies, & Ontologies	3A: Overview
		3B: Applications
		3C: Clinical
Module 4	Data Security, Storage, & Preservation	

ISSN: 2414-1895 DOI: 10.6919/ICJE.201908_5(9).0011

Module 5	Data Sharing & Publishing
Module 6	Data Management Plans

4. Discipline Differences in Data Literacy

Besides emphasizing the establishment of data awareness and the understanding of data knowledge, data literacy is more important to master a series of operable data skill sets, involving the whole scientific research lifecycle from data generation and collection, data analysis and processing, data publication and sharing to data reuse. At the same time, it should be noted that in different disciplines, due to the different knowledge structure, there are different needs for data knowledge and services, so the data attitude and data awareness may be quite different. Therefore, we argue that it is not necessary to emphasize students in all disciplines should have the same level of data literacy. Before setting up, we should pay attention to the focus of the framework in different disciplines, and construct a framework for training students' data literacy ability, which allows disciplines paradigm differences to exist.

From a disciplinary point of view, teachers can define the professional information skills that students need to master, such as how to obtain first-hand information? How to manage large data sets? How to integrate information research into specific research topics? How to grasp the information ethics in the stage of publishing and sharing the research results?

5. Conclusion

The construction of data literacy around scientific research workflow and data lifecycle is of great significance to the implementation of data literacy cultivation. Data literacy will become the decisive factor for college students to acquire knowledge depth and breadth, and the core literacy to adapt to future society and meet challenges.

References

[1] Hao Yuanling, Shen Tingting. An Analysis of the Present situation and upgrade strategies of University Teachers' Data Literacy in the Big Data Era, Journal of Modern Information. Vol.36, No.1, (2016)102-106 (In Chinese)
[2] Qin J, Crowston K, Kirkland A. Pursuing Best Performance in Research Data Management by Using the Capability Maturity Model and Rubrics, Journal of eScience Librarianship. Vol.6, No.2(2017)3
[3] information on https://www.imls.gov/issues/national-initiatives/museums-libraries-and-21st-century-skills/definitions
[4] information on https://library.harvard.edu/services-tools/research-data-management-program
[5] information on https://library.harvard.edu/services-tools/harvard-dataverse
[6] information on https://acrl.ala.org/framework/
[7] information on http://cirss.ischool.illinois.edu/Project/project-details.php?id=19
[8] information on https://nnlm.gov