Research on the Connotation and Structure of Scientific Data Literacy (Postprint)

Authors: Qin Xiaoyan, Chu Jingli

Date: 2023-07-26T00:00:00+00:00

Abstract

Purpose/Significance The conceptual definition of scientific data literacy and an understanding of its intrinsic structure constitute the foundation for conducting theoretical research and educational practice on scientific data literacy. Methods/Process Based on information literacy theory and scientific data lifecycle theory, and building upon domestic and international theoretical research and practical progress in data literacy, this study defines the core connotation of scientific data literacy, analyzes the origin backgrounds and competency requirements of various closely related types of literacy, and establishes an intrinsic structure system for scientific data literacy. Results/Conclusion This study constructs a process structure system based on scientific research workflow and scientific data lifecycle, an objective structure system centered on educational objectives for scientific data literacy in novel data environments, and organically connects and unifies the data behavior process with data cultivation objectives to form a process-objective structure system for scientific data literacy, thereby clarifying the objectives that scientific data literacy should achieve and the specific implementation pathways, laying the theoretical foundation for scientific data literacy research.

Full Text

Preamble

Qin Xiaoyan[1]**, Chu Jingli**[2],[3]
[1] Library of Beihang University, Beijing 100083
[2] National Science Library, Chinese Academy of Sciences, Beijing 100190
[3] University of Chinese Academy of Sciences, Beijing 100049

Abstract

[**Purpose/Significance**] Defining the concept of scientific data literacy and grasping its connotative structure provides the foundation for theoretical research and educational practice in this domain. [**Method/Process**] Based on information literacy theory and scientific data lifecycle theory, this paper defines the core connotation of "scientific data literacy" by building upon domestic and international theoretical research and practical advances in data literacy. It analyzes the origins and competency requirements of various related literacy types and establishes a connotative structure system for scientific data literacy. [**Result/Conclusion**] The study constructs a process structure system grounded in scientific research workflows and data lifecycles, and an objective structure system centered on educational goals for scientific data literacy in new data environments. By organically connecting data behavioral processes with cultivation objectives, a process-objective structure system for scientific data literacy is formed, clarifying both the intended goals and specific implementation pathways. This work lays a theoretical foundation for scientific data literacy research.

Keywords: scientific data literacy; connotation; structure

1. Introduction

Literacy education represents an enduring topic throughout human social development, with definitions and requirements evolving across different historical periods and social contexts. Today, we face explosive data growth and continuous technological advancement, making the ability to effectively identify, acquire, process, evaluate, and utilize data increasingly critical. This capability directly impacts knowledge generation, scientific innovation, and even daily work, learning, and life. Scientific data literacy has emerged as a novel concept in the data-intensive era, and academic analysis of its origins, connotation, and developmental framework constitutes the essential foundation for theoretical research and educational practice [1].

The concept of "information literacy," also termed "information quality," has a long history and mature conceptualization compared to "data literacy." Originating in the late 19th and early 20th centuries, it evolved into "information literacy" by the late 1980s. With the rapid development of information technology and the surge in information volume, the Association of College & Research Libraries (ACRL) issued a landmark report in 1989 that identified information literacy as a survival skill for the information age. The report defined it as "the ability to recognize when information is needed and have the capacity to locate, evaluate, and use effectively the needed information" [2]. This definition has been widely adopted, and the importance of information literacy has been fully recognized. Global efforts to enhance information literacy capabilities have been undertaken by numerous countries and international organizations [3].

The most significant research achievements in information literacy theory primarily encompass two aspects: discussions on the connotative structure of information literacy, and the formation of competency standards and evaluation systems. Regarding structural studies, some scholars propose a "three-level information literacy theory" comprising information knowledge, information capability, and information culture [4], while others argue that information literacy consists of four components: information awareness, information knowledge, information capability, and information ethics [5]. Chinese scholar Pi Jiezheng proposed both process-structure and objective-structure theories of information quality, integrating them to construct a "process-objective structure system of information quality" [6], which provides a theoretical framework for information literacy education.

Regarding competency assessment, numerous institutions worldwide have established corresponding standards and indicator systems, with particularly influential examples including: ACRL's 2000 "Information Literacy Competency Standards for Higher Education" [7] (hereafter "Standards"), the Society of College, National and University Libraries' (SCONUL) 2000 "Seven Pillars of Information Literacy" model and its subsequent revisions [8], the Australian and New Zealand Institute for Information Literacy's (ANZIIL) 2004 "Information Literacy Framework: Principles, Standards and Practice" [9], and UNESCO's 2013 "Global Media and Information Literacy Assessment Framework" [10]. Although these assessment standards have different emphases, they generally establish indicator systems following the information processing and utilization process, adhering to basic workflow patterns of information need identification, acquisition, evaluation, management, and utilization.

ACRL's 2015 "Framework for Information Literacy for Higher Education" [11] (hereafter "Framework") represents a major reform and adjustment. Facing global transformations in higher education and learning environments, along with increasingly complex information ecosystems, the Framework replaced the linear model of learning outcomes and skill lists from the 2000 Standards with an interconnected set of concepts. Through a richer set of core ideas that can be flexibly implemented, it fully taps the potential of information literacy education reform [12]. The Framework extends information literacy to all learning domains for higher education students and integrates it with other academic and social learning objectives. It represents the latest global concept in information literacy evaluation, innovatively introducing two core concepts—"threshold concepts" and "metaliteracy"—that emphasize learners' critical thinking and collaborative abilities regarding information. The Framework's indicators emphasize dynamic, flexible behavioral approaches and critical, deconstructive mindsets [12]. These emerging concepts suggest that scientific data literacy in new research and higher education paradigms should shift from skill cultivation to conceptual enhancement. Existing theoretical achievements in information literacy provide a sufficient theoretical foundation and practical reference for understanding and applying scientific data literacy.

2.2 Scientific Data Lifecycle Theory

The term "lifecycle" originates from life sciences, referring to the entire process a living organism experiences from birth, growth, and maturity to decline and death [13]. This vivid concept, after extension and expansion, has been widely applied to other fields in nature and human society. Lifecycle theory has become an important research method because it examines objects through different stages of their existence, each presenting unique changes and patterns. This stage-based approach to understanding the whole conforms to general laws of development [14]. Information lifecycle theory treats information as an organism with inherent dynamic cycles—from generation and utilization to aging and demise—undergoing continuous processes of value creation, manifestation, and enhancement. The concept of scientific data lifecycle draws upon this theory to study the value decay process of scientific data from creation to archiving and reuse.

Scientific data exists in diverse forms and carriers, and its cyclical process follows inherent patterns, making it suitable for lifecycle research methods. However, scientific data does not completely follow the "value aging" pattern of information resources. Its lifecycle is closely related to scientific research workflows and influenced by disciplinary attributes, research methods, tools, and techniques [15].

Current international theoretical research and practical progress on scientific data lifecycles are advancing rapidly. According to a survey report released by the Committee on Earth Observation Satellites (CEOS) in April 2012, there are already 52 different types of scientific data lifecycle and management models [16]. Several mature and representative models include:

(1) The Inter-university Consortium for Political and Social Research (ICPSR) DDI (Data Documentation Initiative) Combined Lifecycle Model [17] (see [Figure 1: see original paper]). This model comprises six main processes: "Research Concept → Data Collection → Data Processing → Data Distribution → Data Discovery → Data Analysis," with "Data Archiving" and "Data Reuse" as optional paths depending on research needs. This model effectively integrates the data lifecycle with research workflows, clarifying the relationship between data and projects for researchers while establishing a main thread for research data services. U.S. university libraries have built corresponding research data service frameworks based on this model, such as the University of Virginia Library's Research Lifecycle Model [18] and Michigan State University's Records Lifecycle Model [19].

(2) The UK Digital Curation Centre's (DCC) "Data Curation Lifecycle Model" for digital objects or datasets (see [Figure 2: see original paper]). Centered on data, it consists of concentric layers: the first layer is data description and representation; the second is long-term preservation planning; the third involves data sharing and stakeholder participation; the fourth is data management and long-term preservation; and the fifth

encompasses all lifecycle stages related to data curation: "Create or Receive Data → Appraise and Select → Ingest → Preservation Action → Store → Access and Reuse → Transform" [20].

(3) The UK Data Archive (UKDA) Scientific Data Lifecycle, where data undergoes a cyclical structure of six main stages: "Create → Process → Analyze → Preserve → Access → Reuse," with a lifespan longer than the research project itself [21].

(4) The National Science Foundation's (NSF) DataONE (Data Observation Network for Earth) Scientific Data Lifecycle Model, launched in 2009, which identifies eight main stages: "Plan → Collect → Assure → Describe → Preserve → Discover → Integrate → Analyze," forming a cyclical structure [22].

In summary, current scientific data lifecycle models generally unfold around the various stages experienced throughout the entire life process of scientific data. Some studies view the scientific data lifecycle as a linear structure, though more consider it cyclical. Since scientific data is a product of research, it is inseparable from research project workflows. Although natural sciences and humanities/social sciences differ in data migration, transformation, and lifecycle endpoints, certain stages remain common across disciplines, such as data discovery, collection, processing, and analysis. This suggests that when studying the connotative structure of scientific data literacy, we can construct a discipline-general system based on the inherent attributes of scientific data, then supplement it with discipline-specific frameworks reflecting particular research methods and characteristics.

3. Connotation and Characteristics of Scientific Data Literacy

Like many emerging and interdisciplinary fields, data literacy carries different meanings and significance for different stakeholders and experts due to varying historical contexts, academic traditions, and research perspectives. Its connotation in different research contexts cannot be definitively bounded. "Scientific data literacy" belongs to the research domain of "data literacy" while having its own specific conceptual scope, necessitating detailed elaboration and analysis of its connotation and characteristics.

3.1 Connotation of Scientific Data

3.1.1 Data

Traditionally, "data" refers to "justified numbers." Under the dual drivers of informatization and networking, data's connotation and extension have continuously expanded to encompass not only numbers but all information stored on computers, including text, audio, video, etc. The U.S. National Science Foundation (NSF) stated in its 2005 report on "Long-lived Digital Data Collections" that data refers to any information that can be stored digitally, including text, numbers, images, video or film, audio, software, algorithms, animations, models, simulations, etc. [23]. China's national standard (GB/T18391.1-2009) defines "data" as "a formalized representation of facts, concepts, or instructions suitable for communication, interpretation, or processing by humans or automated means" [24]. Data can represent many different things and has numerous classification methods; by generation method, data can be categorized as observational, experimental, simulated, derived or compiled, and reference or normative [25].

With the development of big data, cloud computing, mobile internet, IoT, and artificial intelligence, all content, media, or publications can be viewed as data collections from a digital and computational perspective. These data objects can be digitally, structurally, and semantically parsed, annotated, and linked to become logically meaningful and interrelated knowledge objects. All these knowledge objects are computable, recombinable, integrable, and recreatable, and can interact with users in personalized and dynamic ways to form new data objects and knowledge content. Various data objects have become explicit, fundamental, and primary knowledge resources, marking our transition from an information environment to a data environment [26].

In today's era, the concept of data has far exceeded existing definitions. Due to its diverse forms and rich connotations, data has essentially evolved into a collective term for information, intelligence, knowledge, and semantics. The academic communication system has consequently undergone significant changes, with data serving as both an important foundational resource for research and an output form of scientific achievements, playing a crucial role in the scientific research process.

3.1.2 Scientific Data

What constitutes scientific data? Major foreign terms include "Science Data," "Scientific Data," or "Research Data," while domestically it has evolved from "scientific and technological data" to "scientific data" [27]. Definitions from domestic and international research institutions and scholars include:

(1) **Traditional scientific data**: China's Scientific Data Sharing Engineering defines scientific data as original and foundational data generated by humans in scientific and technological activities to understand and transform the world, as well as systematically processed data products and related information. This includes massive long-term data accumulated and compiled through large-scale observation, detection, investigation, experimentation, and comprehensive analysis by public welfare sectors, as well as substantial data generated through national science and technology programs and years of scientific practice by researchers [28]. The OECD [29], U.S. National Institutes of Health [30], Australian National University [31], and University of Cambridge [32] have similarly defined "research data" from related perspectives. For domain-specific scientific data, many institutions have provided interpretations, such as China's Agricultural

Scientific Data Sharing Center, which defines agricultural scientific data as basic data generated from agricultural scientific and technological activities and systematically processed data products and related information [33].

Synthesizing these research findings, institutions have defined scientific data from a management perspective, roughly falling into three categories: first, actual recorded materials that can verify research findings, excluding unverified data such as notes and preliminary experimental results; second, semi-finished products including experimental data, notes, images, audio/video, and simulation systems related to scientific research; and third, the broadest understanding encompassing all process data, semi-finished products, and research outcomes throughout the entire scientific research process.

(2) **Scientific data in the digital research context**: With advances in information technology, scientific research informatization has matured, and scientific computing has enhanced humanity's ability to describe complex social phenomena. Various network transmission systems and data storage and analysis facilities not only provide scientists with powerful observation, analysis, and experimental capabilities but also promote the evolution of research methods from theoretical analysis and observation to simulation of research objects, driving the formation of "the fourth paradigm of data-intensive scientific research" and a shift from hypothesis-driven to data-driven scientific methods [34]. This paradigm shift has endowed scientific data with new meanings:

First, large volumes of raw scientific data are generated through collection by detectors and other high-end equipment and high-performance computer simulations. This includes not only theoretical predictions and experimental observations from traditional conditions but also comprehensive factors from researchers, scientific instruments, research processes, and management mechanisms, as well as digital expressions produced through computer simulation and modeling analysis, including research objects in social science fields. The advantage of this digital expression lies in its ability to describe large-scale or micro-scale entities and enable various forms of combination, transformation, and digital expression according to research needs [15].

Second, initial data includes unprocessed scientific experimental data and digital expressions of research objects. Scientific workflow technologies from initial to intermediate data and final results enable visualization methods, techniques, and software resources for scientific data to exist as a resource type, allowing people to describe scientific research or experimental processes in a reproducible, verifiable, and distributed manner, understand data processing methods, models, and tools employed, and thereby verify scientific research [35].

Additionally, in the big data era, "data" is no longer isolated and static but dynamic and systematic, existing as continuous and uninterrupted "data flows"—vast, interrelated, and living data. Data has become an essential foundational resource for innovation across all social sectors, making systematic collection and scientific analysis crucial.

(3) **Scientific data studied in this paper**: Given the evolution from traditional research environments to e-research and big data eras, the generation pathways, organizational forms, and application domains of scientific data have undergone tremendous changes. This paper defines scientific data as digital information suitable for communication, interpretation, or processing—generated through observation, experimentation, investigation, recording, or computation in natural sciences, engineering technology, and humanities/social sciences—as well as systematically processed or reorganized data products and related information according to different research needs. This includes numerical data, textual data, images, audio, video, and other multimedia data usable for scientific research, technical design, verification, and decision-making.

3.2 Connotation of Scientific Data Literacy

"Data literacy" originated in American educational circles, with its connotation evolving alongside technological development, showing variations across different eras and fields. No standardized definition has yet formed in existing domestic and international research. Terminological expressions include: "Data Literacy," "Data Information Literacy," "Science Data Literacy," "Research Data Literacy," and "Data Management Literacy."

The proposal of "scientific data literacy" confines the concept and application scope of "data literacy" to the scientific research domain and provides clearer boundaries for "data," specifically referring to original and foundational data generated in scientific and technological activities to understand and transform the world, as well as systematically processed data products and related information. "Literacy" is defined as "the ability to read and write" or "an accomplishment acquired through training or practice," encompassing moral character, appearance, knowledge level, and capabilities. From this definition, "scientific data literacy" can be understood as "the thinking, knowledge, and skills involved in collecting, processing, managing, evaluating, and utilizing data for scientific research, as well as the ethical norms and behavioral standards universally followed throughout the data lifecycle."

Scientific data literacy emphasizes understanding, utilizing, and managing scientific data to transform data into knowledge. Rapid development of information technology and digital information resources has revolutionized scientific research activities. In e-science and e-social science, as well as data-intensive research paradigms, new research environments demand enhanced data utilization capabilities from researchers, while data-intensive paradigms significantly increase the need for data management skills and related professionals. Researchers must not only effectively analyze and utilize scientific data but also systematically collect and manage it.

Existing scientific data literacy research generally follows two paths: first, a data utilization perspective focusing on user data behavior research within the information literacy domain—namely, how to use and reuse scientific data; second, a data management perspective focusing on data production, organization, and storage throughout the research lifecycle. These two objectives are both distinct and interrelated, fundamentally aiming to cultivate learners' knowledge and skills in collecting, processing, managing, evaluating, and using data during scientific exploration, enabling them to possess scientific data literacy capabilities required for research innovation and development.

3.3 Relationship Between Scientific Data Literacy and Other Literacies

"Literacy" is a concept that continuously develops and enriches with changing times. When a behavior or lifestyle becomes increasingly popular and influential, traditional literacy's role and value gradually weaken, objectively necessitating new literacy concepts that enable individuals to realize their knowledge and potential, fully participate in their communities and broader society, adapt to challenges and requirements of social development, and achieve their goals. "Scientific data literacy" has emerged based on related literacy concepts such as information literacy, digital literacy, scientific literacy, and statistical literacy. Clarifying its connections and distinctions with these related literacies is prerequisite to understanding scientific data literacy.

3.3.1 Relationship with Information Literacy

Since its emergence in 1974, information literacy has sparked worldwide research and discussion. The most influential definition comes from ACRL's 2000 "Information Literacy Competency Standards for Higher Education," which cited the American Library Association's (ALA) authoritative 1989 definition: "the ability to recognize when information is needed and have the capacity to locate, evaluate, and use effectively the needed information" [7]. Today, information technology and environments have transformed dramatically, and information literacy's connotation has been elaborated from multiple perspectives. ACRL's 2015 "Framework for Information Literacy for Higher Education" enriched the concept to "the set of integrated abilities encompassing the reflective discovery of information, the understanding of how information is produced and valued, and the use of information in creating new knowledge and participating ethically in communities of learning" [11], emphasizing dynamism, flexibility, and integration with emerging literacies.

In the big data era, scientific data literacy centered on data management and utilization should become an important component of information literacy in higher education. To some extent, scientific data literacy can be viewed as an extension and expansion of information literacy, as research institutions and experts suggest that research data should be treated as a type of information, and data management knowledge and skills should be incorporated into information literacy definitions to ensure students acquire essential career skills [36-39].

However, scientific data literacy differs fundamentally from traditional information literacy in practice, primarily because data complexity far exceeds that of other information types. While information literacy reflects the ability to find, acquire, evaluate, and use information, scientific data literacy reflects the ability to understand, use, and manage scientific data, focusing on data's special attributes and application value.

3.3.2 Relationship with Digital Literacy

"Digital literacy" was formally proposed by P. Gilster in 1997, defined as "the ability to understand, evaluate, and use various digital resources and information based on computer technology" [40], with critical thinking as its core rather than merely digital technical skills. Israeli scholar Y.E. Eshet-Alkalai constructed a five-framework concept of digital literacy based on years of research: photo-visual literacy, reproduction literacy, branching literacy, information literacy, and socio-emotional literacy, representing respectively the abilities to understand graphical information, reintegrate information, conduct nonlinear information searches, construct knowledge from fragmented information, retrieve/filter/discern/use information, and share knowledge through digital emotional communication [41]. IFLA's 2017 digital literacy declaration states that "being digitally literate means being able to maximize the use of digital technology to meet personal, social, and professional information needs in an efficient, effective, and ethical manner" [42].

Digital literacy emphasizes mastering and applying digital technology and network knowledge to identify, organize, understand, evaluate, and create information to meet various learning, working, and living needs—a lifelong learning process. When emphasizing scientific data literacy capabilities as data producers, users, and managers, we cannot separate from digital literacy's basic requirements. However, given data's rich types and complex processing workflows, scientific data literacy's research scope is broader and more specialized.

3.3.3 Relationship with Scientific Literacy

International consensus generally summarizes scientific literacy as three components: understanding scientific knowledge, mastering scientific research processes and methods, and understanding the impact of science and technology on society and individuals [43-44]. Scientific literacy represents the basic quality and cultivation researchers demonstrate during scientific research, fundamental to successful research innovation and task completion. This means everyone engaged in scientific research should possess scientific literacy, which involves the entire scientific thinking system, values, cultural identity, and scientific communication patterns, directly relating to reading, understanding, and producing research outcomes [45]. Scientific data literacy is closely related to scientific literacy, both involving methods, approaches, attitudes, and skills related to scientific thinking and research, with scientific data literacy enriching scientific literacy's connotation.

3.3.4 Relationship with Statistical Literacy

"Statistical literacy" refers to the ability to understand and reasonably use statistical data and methods to solve practical problems and make decisions based on comprehensive information [46]. China's compulsory education mathematics curriculum standards provide a comprehensive explanation: "the ability to think about data-related issues from a statistical perspective; make reasonable decisions through data collection, description, and analysis; recognize statistics' role in decision-making; and reasonably question data sources, processing methods, and results" [47]. Statistical literacy's core is critical thinking about and application of data, with computational ability as an important factor.

In the big data era, statistical literacy—understanding and utilizing statistical information including methods—has strong relevance to data science, and traditional boundaries between data science and statistics have become increasingly blurred. Scholars propose that as data's role and nature change, statistical literacy's traditional definition should be updated to reflect data's enormous role in our lives, with data literacy enriching statistical literacy's connotation [48].

4. Connotative Structure of Scientific Data Literacy

What specific elements should scientific data literacy include? What are their internal relationships and structure? How can we systematically and clearly depict its connotative structure? Analysis reveals that after defining scientific data literacy, its application generally unfolds according to scientific data utilization behaviors or data problem-solving processes. This process maintains high stability in both content and order, comprising multiple logical stages. Requirements for data actors in cognition, skills, and affect at each stage constitute scientific data literacy's connotation. Combining research workflows and scientific data lifecycles, this paper constructs a process structure system for scientific data literacy (see [Figure 3: see original paper]).

The process structure shows that in digital research and big data eras, data-centered behavioral processes can be condensed into: data needs analysis, data production and collection, data analysis and processing, data publication and sharing, data organization and preservation, data discovery and acquisition, and data evaluation and reuse. This entire process develops vertically, with each stage imposing corresponding knowledge and skill requirements on actors, as well as specific demands on awareness, concepts, and ethics—such as knowledge about data sources and environments, mastery of data processing technologies, data acquisition and evaluation capabilities, and moral and normative principles to follow throughout. Actors' attitudes and affect also influence the entire process.

4.2 Objective Structure of Scientific Data Literacy

Beyond understanding scientific data literacy from the data behavior process itself, some research, particularly by domestic scholars, abstracts educational objectives for scientific data literacy—what target requirements learners should meet under new big data environments to be considered scientifically data-literate [49-52]. Although researchers have different understandings and expressions of component concepts, essential commonalities exist: all include data knowledge, data skills, and data cognition, forming a three-layer structure with data skills as the hub connecting data knowledge and data cognition. Data knowledge forms the foundation for all data-related activities. Data skills emerge from knowledge accumulation and application, and their development enriches knowledge while deepening data cognition. Data cognition 主要包括两方面的内容，一方面是人们对数据价值的认识和研究兴趣，即数据意识和数据文化，另一方面是认识到数据发展中的各种行为规范和道德准则，即数据伦理，保证数据主体的数据行为遵循正确的方向，从而维护数据生态的正常秩序，确保人们在数据活动中的利益和社会整体利益的一致性。

Through consolidation and refinement of major perspectives, five interrelated, mutually foundational, and closely connected components can be extracted to form the complete connotation of scientific data literacy: data awareness, data culture, data knowledge, data skills, and data ethics. These five components should become the target connotation when designing and implementing scientific data literacy competency indicator systems and educational activities. This connotative structure theory is termed the objective structure system of scientific data literacy (see [Figure 4: see original paper]).

4.3 Process-Objective Structure

Scientific data literacy encompasses commonalities from information literacy, digital literacy, scientific literacy, and statistical literacy while possessing unique distinguishing characteristics. Its complexity makes conceptual understanding challenging, yet accurate comprehension is essential for establishing evaluation standards and educational content. This study deeply analyzes domestic and foreign scholars' shared starting points in researching related literacy connotations: focusing on key roles in the data ecosystem—(1) individuals, whose data behavioral processes have specific origins and histories that should constitute scientific data literacy's inherent structure; and (2) society, whose objective requirements shape educational goals.

Although scholars' methods and final connotative structures differ in direction and expression, they share unifying characteristics. In scientific data literacy research, these common features can be connected and unified to form a process-objective structure system (see [Figure 5: see original paper]), where requirements at each process stage can be summarized as objectives, and objectives' concrete manifestations are behavioral processes. Objectives, through cognitive, behavioral, and affective elevation and internalization, ultimately form scientific data literacy.

The process-objective structure system provides a more comprehensive and complete explanation of scientific data literacy's connotation and structure, reflecting its essential meaning and practical value. It enhances the operability of educational practice and facilitates evaluation and implementation. This dynamic theoretical framework will continuously self-adjust and develop with the data era, forming a scientific, reasonable, dynamic, and stable scientific data literacy system.

Conclusion

Defining scientific data literacy and grasping its connotative structure establishes the theoretical foundation for educational frameworks. Based on information literacy theory, scientific data lifecycle theory, and domestic/international research advances, this paper defines the specific scope of "scientific data" and determines the connotation and characteristics of "scientific data literacy." Through analysis of information literacy, digital literacy, scientific literacy, and statistical literacy—their origins and competency requirements—this study deepens understanding of scientific data literacy's core connotation and research scope. The constructed connotative structure system includes a process structure based on research workflows and data lifecycles, and an objective structure centered on educational goals in new data environments. By organically connecting data behavioral processes with cultivation objectives, the process-objective structure system clarifies both the goals and implementation pathways, laying a theoretical foundation for scientific data literacy research.

References

[1] Qin Xiaoyan. Research on the Construction and Empirical Study of Scientific Data Literacy Capability Indicator System[D]. Beijing: University of Chinese Academy of Sciences, 2018.
[2] Association of College and Research Libraries. Presidential Committee on Information Literacy: Final Report[R/OL]. 1989[2018-01-10]. http://www.ala.org/acrl/publications/whitepapers/presidential.
[3] Horton F W. Understanding Information Literacy: A Primer. Information Society Division, Communication and Information Sector[EB/OL]. [2018-01-10]. http://unesdoc.unesco.org/images/0015/001570/157020e.pdf.
[4] Qin Dianqi, Zhang Yuwei. Construction and Practice of Three-Level Information Literacy Theory[J]. Information Studies: Theory & Application, 2017, 40(6): 13-17.
[5] Han Jingxian, Zhao Manjuan. Information Literacy Education: Theory and Practice[M]. Guangzhou: World Book Guangdong Publishing Company, 2014: 1.
[6] Pi Jiezheng. Research on Information Quality Theory and Education[D]. Beijing: Graduate School of Chinese Academy of Sciences (National Science Library), 2003.
[7] Association of College and Research Libraries. Information Literacy Competency Standards for Higher Education[R/OL]. [2018-02-01]. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf.
[8] Society of College, National and University Libraries. The Seven Pillars of Information Literacy[EB/OL]. [2018-02-01]. https://www.sconul.ac.uk/page/seven-pillars-of-information-literacy.

[9] Bundy A. Australian and New Zealand Information Literacy Framework: Principles, Standards and Practice[M]. 2nd ed. Adelaide, SA, Australia: Australian and New Zealand Institute for Information Literacy, 2004.
[10] UNESCO. Global Media and Information Literacy Assessment Framework: Country Readiness and Competencies[EB/OL]. [2018-08-10]. http://unesdoc.unesco.org/images/0022/002246/224655e.pdf.
[11] Association of College and Research Libraries. Framework for Information Literacy for Higher Education[R/OL]. [2018-08-10]. http://www.ala.org/acrl/standards/ilframework.
[12] Qin Xiaoyan. Improvement and Enlightenment of American University Information Literacy Standards: Interpretation of ACRL's Framework for Information Literacy for Higher Education[J]. Library and Information Service, 2015, 59(19): 139-144.
[13] Zhu Xiaofeng. Lifecycle Methodology[J]. Studies in Science of Science, 2004(6): 566-571.
[14] Ma Feicheng, Xia Yonghong. Empirical Research on Network Information Lifecycle[J]. Information Studies: Theory & Application, 2009, 32(6): 1-7.
[15] Liu Xia. University Scientific Data Management and Empirical Research[M]. Wuhan: Wuhan University of Technology Press, 2013.
[16] CEOS. Data Lifecycle Models and Concepts[EB/OL]. [2018-02-10]. http://wgiss.ceos.org/dsig/whitepapers/.
[17] DDI Structural Reform Group. Overview of the DDI Version 3.0 Conceptual Model[EB/OL]. [2018-02-12]. http://opendatafoundation.org/ddi/srg/Papers/DDIModel_{v_4}.pdf.
[18] University of Virginia Library. Research Data Services[EB/OL]. [2018-11-06]. http://data.library.virginia.edu/data-management/.
[19] Michigan State University. Records Management[EB/OL]. [2018-02-12]. http://archives.msu.edu/records/.
[20] Higgins S. The DCC Curation Lifecycle Model[J]. International Journal of Digital Curation, 2008, 3(1): 134-140.
[21] UK Data Archive. Research Data Lifecycle[EB/OL]. [2018-02-12]. http://www.data-archive.ac.uk/create-manage/life-cycle.
[22] Michener W K, Jones M B. Ecoinformatics: Supporting Ecology as a Data-Intensive Science[J]. Trends in Ecology & Evolution, 2012, 27(2): 85-93.
[23] NSF. National Science Board. Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century[R/OL]. [2018-03-26]. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf.
[24] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. Information Technology—Metadata Registries (MDR) Part 1: Framework. GB/T18391.1-2009[S]. Beijing: Standards Press of China, 2009.
[25] University of Virginia Library. Data Types & File Formats[EB/OL]. [2018-03-26]. http://data.library.virginia.edu/data-management/plan/format-types/.
[26] Zhang Xiaolin. Disruptive Changes and the Post-Library Era: Promoting Supply-Side Structural Reform of Knowledge Services[J]. Journal of Library Science in China, 2018(1): 1-12.
[27] Sun Jiulin, Huang Dingcheng, Li Xiaobo. New Progress in China's Scientific and Technological Data Management and Sharing Services[J]. World

Sci-Tech R&D, 2002(5): 15-19.
[28] Scientific Data Sharing Engineering Technical Standard SDS/T1003.2-2004. Scientific Data Sharing Concepts and Terms[S/OL]. [2018-03-11]. http://www.sciencedata.cn/pdf/2.pdf.
[29] OECD. OECD Principles and Guidelines for Access to Research Data from Public Funding[R/OL]. [2018-05-16]. http://www.oecd.org/dataoecd/9/61/38500813.pdf.
[30] National Institutes of Health. NIH Grants Policy Statement[EB/OL]. [2018-03-17]. http://grants.nih.gov/grants/policy/nihgps_{2011}/nihgps_{ch2}.htm.
[31] The Australian National University. Information Literacy Program. ANU Data Management Manual: Managing Digital Research Data at the Australian National University[EB/OL]. [2018-01-20]. https://services.anu.edu.au/files/DataManagment.pdf.
[32] University of Cambridge. Explanation of Terms[EB/OL]. [2018-03-27]. http://www.lib.cam.ac.uk/preservation/incremental/glossary.html.
[33] Introduction to Agricultural Scientific Data Sharing Center[EB/OL]. [2018-01-15]. http://www.agridata.cn/homepage/ch_{intro}.asp.
[34] Hey T, Tansley S, Tolle K. The Fourth Paradigm: Data-Intensive Scientific Discovery[M]. Translated by Pan Jiaofeng, Zhang Xiaolin. Beijing: Science Press, 2012.
[35] Qian Peng. Research on University Scientific Data Management[D]. Nanjing: Nanjing University, 2012.
[36] Andretta S, Pope A, Walton G. Information Literacy Education in the UK: Reflections on Perspectives and Practical Approaches of Curricula[J]. Communications in Information Literacy, 2008, 2(1): 36-51.
[37] Research Information Network. The Role of Research Supervisors in Information Literacy[R/OL]. [2018-10-22]. http://www.rin.ac.uk/system/files/attachments/Summary_{{{of}}_{{
[38] Schneider R. Research Data Literacy[C]//European Conference on Information Literacy. Springer International Publishing, 2013: 134-140.
[39] Vitae. Researcher Development Framework[R/OL]. [2018-05-08]. https://www.tes.ac.uk/docs/DocRepo/Research/Vitae-Researcher-Development-Framework.pdf.
[40] Gilster P. Digital Literacy[M]. New York: Wiley, 1997: 25-28.
[41] Eshet-Alkalai Y. Digital Literacy: A Conceptual Framework for Survival Skills in the Digital Era[J]. Journal of Educational Multimedia and Hypermedia, 2004, 13(1): 93-106.
[42] He Lei. IFLA Digital Literacy Declaration[J]. Library Tribune, 2017, 37(11): 1-4.
[43] Shamos M H. The Myth of Scientific Literacy[M]. New Brunswick, NJ: Rutgers University Press, 1995.
[44] Jenkins E W. Scientific Literacy: A Functional Construct[M]//Baker D, Clay J, Fox C. Challenging Ways of Knowing in English, Maths, and Science. London: Falmer Press, 1996: 43-51.
[45] Elmborg J. Critical Information Literacy: Implications for Instructional Practice[J]. Journal of Academic Librarianship, 2006, 32(3): 192-199.
[46] Li Jun. On the Cultivation of Statistical Literacy[J]. Journal of Zhejiang Education Institute, 2009(01): 10-16.
[47] Ministry of Education of the People's Republic of China. Full-Time

Compulsory Education Mathematics Curriculum Standards: Experimental Draft[M]. Beijing: Beijing Normal University Press, 2001.
[48] Gould R. Data Literacy is Statistical Literacy[J]. Statistics Education Research Journal, 2017, 16(1): 22-25.
[49] Zhang Xiaoyang, Li Mei. Research on a Competency-Based Assessment Scale for Graduate Students' Data Literacy[J]. Library and Information Service, 2017, 61(8): 89-95.
[50] Zhang Bin, Liu Sannüya, Liu Zhi, et al. Research on Cultivation Strategies for Normal University Students' Data Literacy Based on Big Data[J]. e-Education Research, 2017, 38(12): 86-91, 120.
[51] Ling Wanyang. Research on Researchers' Data Literacy Under Big Data and Data-Intensive Research Paradigms[J]. Library, 2018(1): 81-87.
[52] Bu Binghua. Review of Research on College Students' Data Literacy Education in China Under Big Data Environment[J]. Digital Library Forum, 2017(12): 63-67.

Author Contributions

Qin Xiaoyan: Data investigation and paper writing.
Chu Jingli: Paper revision and final approval.

Research on the Connotation Structure of Scientific Data Literacy

Qin Xiaoyan[1], Chu Jingli[2],[3]
[1] Library of Beihang University, Beijing 100083, China
[2] National Science Library, Chinese Academy of Sciences, Beijing 100190, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: [Purpose/significance] The definition of scientific data literacy and understanding of its connotative structure form the basis for theoretical research and educational practice. [Method/process] Based on information literacy theory and scientific data lifecycle theory, this paper defines the core connotation of "scientific data literacy" by examining domestic and international research and practical progress. It analyzes the origins and competency requirements of closely related literacy types and establishes a connotative structure system. [Result/conclusion] The study constructs a process structure system based on research workflows and data lifecycles, and an objective structure system centered on educational goals in new data environments. By organically connecting data behavioral processes with cultivation objectives, a process-objective structure system is formed, clarifying the goals and implementation pathways for scientific data literacy, thereby laying a theoretical foundation for related research.

Keywords: scientific data literacy; connotation; structure

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