Constructing a Scientific Data Quality Assessment Framework Based on Lifecycle Management Stages of Scientific Data: Postprint

Authors: Jiang Hong, Wang Chunxiao

Date: 2023-04-01T16:15:54+00:00

Abstract

[Objective/Significance] By selecting scientific data quality evaluation indicators from 15 domestic and international scientific data centers, this study aims to identify common indicators that can objectively reflect scientific data quality and construct a universal scientific data quality evaluation indicator system. [Method/Process] Through document survey method, web survey method, and content analysis method, this research systematically reviews and analyzes the scientific data evaluation indicators of the 15 scientific data centers to gain insights into the data evaluation indicators of existing scientific data institutions. [Results/Conclusions] Based on the various stages of scientific data lifecycle management, a scientific data quality evaluation indicator model is constructed, comprising five dimensions: data management plan, data collection management, data analysis and processing management, data preservation management, and data sharing and utilization management. This provides a reference for establishing decision-oriented evaluation systems for national and local scientific data centers in China.

Full Text

Research on the Construction of Scientific Data Quality Evaluation System Based on Scientific Data Lifecycle Management Phases

Jiang Hong[1], Wang Chunxiao[1,2]
[1]Wuhan Library, Chinese Academy of Sciences, Wuhan 430071
[2]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract:
[Purpose/Significance] This study selects scientific data quality evaluation indicators from 15 scientific data centers at home and abroad, aiming to screen common indicators that can objectively reflect scientific data quality and construct a universal scientific data quality evaluation index system. [Method/Process] Using document investigation, web survey, and content analysis methods, the scientific data evaluation indicators of 15 scientific data centers were sorted and analyzed to understand the existing data evaluation indicators of scientific data institutions. [Result/Conclusion] Based on the various stages of scientific data lifecycle management, a scientific data quality evaluation index model composed of five dimensions—data management plan, data collection management, data analysis and processing management, data storage management, and data sharing and utilization management—was constructed to provide reference for the establishment of decision-oriented scientific data center evaluation systems for national and local scientific data centers in China.

Keywords: scientific data; data lifecycle; evaluation system; indicators
Classification Number: G203
DOI: 10.13266/j.issn.0252-3116.2020.10.003

With the continuous development of scientific research, the volume of scientific data has become increasingly massive and its structure increasingly complex. Scientific data holds tremendous research value, and the study of scientific data has become a top priority in scientific research. The evaluation of scientific data is a critical link that both scientific data management and service institutions must emphasize. Foreign scientific data institutions have conducted substantial work in evaluation, and some of their experiences are worth learning from for domestic scientific data centers. For example, the information quality guidelines of the U.S. National Oceanic and Atmospheric Administration (NOAA) propose indicators such as usability, objectivity, integrity, impact, transparency, and reproducibility, along with their application scopes [1]. The Dutch Data Archiving and Networked Services (DANS) evaluates datasets on its online storage system using indicators of findability, accessibility, interoperability, and reusability [2], which also constitute the FAIR principles in scientific data management [3]. China began requiring the submission of scientific data generated by national projects in 2008 and subsequently issued various scientific data submission measures for different projects [4]. The "Scientific Data Management Measures" issued by the State Council in 2018 further demonstrated the nation's emphasis on scientific data as a strategic resource. Establishing a reasonable scientific data quality evaluation index system can promote the construction of national and local scientific data centers in China and facilitate the gradual establishment of renowned scientific data evaluation research centers. However, current domestic research on scientific data evaluation remains limited, and scientific data evaluation lacks unified standards. Through investigating the scientific data quality evaluation indicators of scientific data institutions at home and abroad, this paper attempts to establish a scientific data quality evaluation model from the perspective of scientific data lifecycle management, hoping to provide reference for relevant research and work on scientific data evaluation in China.

1. Research Status on Scientific Data Quality Evaluation

1.1 International Research Status

From the perspective of broad data quality, foreign scholars have paid considerable attention to data quality evaluation and constructed numerous data quality evaluation models. For instance, B. Stvilia et al. constructed an information quality evaluation model from three dimensions: intrinsic information quality, contextual information quality, and reputation information quality [5]. C. Batini et al. proposed methodology-based data quality dimensions including completeness, accuracy, timeliness, consistency, accessibility, credibility, usability, interpretability, and appropriate data volume [6]. A. Zaveri et al. constructed 18 different data quality dimensions to evaluate linked data, grouping these dimensions into four categories: accessibility, contextual, intrinsic, and representational dimensions [7]. Some scholars have constructed scientific data quality evaluation models for specific disciplinary fields: M. G. Kahn et al. built a quality evaluation model for clinical research data from electronic health records, with main indicators including accuracy, credibility, objectivity, timeliness, and reasonableness of data volume [8]; H. Chen et al. constructed a three-dimensional data quality evaluation model to assess public health-related data, with the three dimensions being the data itself, data usage, and data collection process [9]; H. Huang et al. proposed data quality standards applicable in genome annotation environments based on previous data quality evaluation criteria [10]. The investigation reveals that although the data quality dimensions proposed by different scholars are numerous, they contain overlapping and repetitive content. When evaluating specific scientific data, dimensions have different priorities in different usage contexts. At present, scholars mostly focus on evaluation research for data or technical platforms in specific disciplinary fields.

1.2 Domestic Research Status

Domestic research has included numerous studies on data evaluation, primarily concerning government open data evaluation. For example, Shao Yanhong constructed an evaluation standard for government open data quality based on existing evaluation indicators and data quality standards [11]. Li Xiaotong et al., based on a survey of quality issues in over 1,900 datasets from Beijing, Guangzhou, and Harbin, extracted seven quality dimensions and measurable evaluation indicators: completeness, timeliness, consistency, accuracy, uniqueness, understandability, and openness [12]. Many scholars have focused on the evaluation of scientific data platform construction. Liu Guifeng et al. analyzed six international organization projects on open government data, extracted indicators applicable to scientific data platforms, and constructed a scientific data platform evaluation index system from four aspects: platform infrastructure, platform management functions, platform data, and platform effects and impacts, incorporating data lifecycle theory into secondary indicators [13]. Zhou Yu et al., through investigating data curation platforms at home and abroad and using expert surveys, ultimately determined an evaluation index system for data curation platforms, including dimensions such as data management systems, service functions, data volume, data quality, platform interface, software systems, and utilization rate [14]. In addition to platform evaluation, some researchers have focused on data quality itself. Yu Fangdong constructed an indicator framework from three aspects—data source conditions, metadata, and data quality—to evaluate non-traditional data in government statistics [15]. Yu Houqiang et al. constructed an altmetric data quality evaluation system by sorting out the altmetric data production process [16]. Domestic research on scientific data quality evaluation remains limited. Currently, evaluation systems issued by domestic scientific data institutions almost only involve broad indicators such as accuracy, completeness, and usability. This paper attempts to construct a quality evaluation model applicable to different stages of scientific data management lifecycle by investigating scientific data institutions across various disciplinary fields.

2. Research Methods

Ma Feicheng and Wang Juncheng argue that objects suitable for lifecycle methods should possess three important attributes—"continuity, irreversibility, and iterativity." Different stages of the life process are not only continuous but also temporally irreversible; after completing one life cycle, they enter the next cycle, and the transition between cycles represents iteration or circulation [17]. Based on this theory, Ding Ning et al. propose that lifecycle methods can also be applied to scientific data. The scientific data lifecycle is closely related to the scientific research process, and the essence of scientific data lifecycle management is managing data according to scientific research workflows [18]. Different research activities may only include some stages of the scientific data lifecycle; for example, a research project primarily focused on data processing and analysis might bypass data generation and collection stages [19]. Analyzing data evaluation indicators from the perspective of scientific data lifecycle management can more clearly reflect the characteristics of scientific data management based on research processes, facilitating more effective scientific data lifecycle management. Zhang Yang and Xiao Yanzhu, through investigating and analyzing ten data lifecycle theories, summarized five core stages of scientific data lifecycle management: data management plan formulation, data collection management, data analysis and processing management, data storage management, and data sharing and utilization management [20]. This paper uses these five stages of scientific data lifecycle management as dimensions to analyze specific evaluation indicators for scientific data.

This study employs document investigation, web survey, and content analysis methods. Based on extensive web surveys, 15 data institutions with explicitly proposed indicators were selected (see Table 1), primarily distributed in the United States, Europe, and China. During the investigation, the author focused on the evaluation indicators of these data institutions across the five stages of scientific data management plan, data collection management, data analysis and processing management, data storage management, and data sharing and utilization management. The scientific data content of these 15 institutions covers geography, biology, medicine and health, society, economy, and other natural science fields, providing relatively comprehensive coverage. Some data institutions concentrate their data resources in one or two disciplinary fields, while others focus on numerous disciplines with rich data resources.

3. Analysis of Scientific Data Quality Evaluation Indicators

3.1 Data Management Plan Formulation

The "Scientific Data Management Measures" emphasizes that legal entities and competent departments at all levels should formulate scientific data management plans and fulfill their responsibilities in scientific data management [21]. The main task of this stage is to plan how to describe and store data, which involves having complete metadata standards (e.g., defining data types and formats) and determining how to manage, access, and share data throughout the data lifecycle, such as assigning responsibilities for data management and ensuring qualified professionals execute the data management plan. The data management plan reflects the awareness and capabilities of researchers and research organizations regarding data management. Analysis of the investigation results shows that indicators in this stage mainly relate to four aspects of data management plan (DMP) formulation: DMP completeness. ORNL DAAC emphasizes that the provided DMP should include as much content as needed to describe the data, such as reasonable explanations of precision and density for specific data types [22]; SSDA believes that comprehensive data documentation explaining how data were created should be developed during the planning stage [23]; ARM stipulates that DMPs must describe how data will be shared and preserved and include requirements regarding personal privacy and confidential information [24]; EROS considers the process integrity of data management and budget expenditures [25]. Data management responsibilities. EROS emphasizes that there should be professional personnel for data management at this stage; the scope of management responsibilities should be clearly defined; requirements should comply with official institutional standards; data documentation including metadata development and maintenance should be ensured; and data quality standards should be established [26]. DMP value. Both SSDA and ARM emphasize DMP value. The former believes DMPs should be convincing and gain support from funders [27]; the latter believes DMPs should have value in facilitating research [24]. DMP formulation should be convenient and easy to operate. NCL Data emphasizes providing multiple DMP formats for reference; providing training, guidelines, or assistance for DMP formulation; providing links to DMP creation tools; and offering contact information for consultation during DMP creation [28].

Evaluating the DMP itself during the planning stage is an evolving new concept [29]. A good DMP document should explain matters needing attention throughout all stages of the data lifecycle and reflect the requirements of specific research projects and funding agencies. The investigated institutions provided limited explanation of this indicator.

3.2 Data Collection Management

Scientific data are increasingly recognized as valuable by scientific and technological departments and researchers at all levels in China, serving as an important foundation for new rounds of scientific and technological innovation. The construction of scientific data centers cannot be separated from scientific data collection. The "Measures" emphasize that legal entities should undertake the integration and submission of scientific data in their relevant fields, and each unit should have a scientific data quality control system to ensure the accuracy and usability of data expression [21]. Investigation and analysis of this stage reveal that evaluation indicators for scientific data mainly include five aspects: (1) Data collection format requirements. Deep Blue Data and ORNL DAAC recommend submitting data in non-proprietary, open formats [22,30]; UniProtKB recommends using data formats that comply with UniProtKB requirements [31]; DRUM believes submitted data should conform to given file formats suitable for access, with different format regulations for different data types [32]; TDR has format requirements for tabular data files, requiring SPSS (POR and SAV formats), STATA, Rdata, CSV, and other formats [33]; NGDC also emphasizes that data submission should adopt prescribed standard formats [34]. (2) Data review. CNGDb indicates that submitted data need to pass MD5 verification for data transmission integrity and must undergo review of metadata information and ethical approval documents [35-36]; ORNL DAAC emphasizes reviewing data priority areas, scientific impact, and community needs to determine data priority [37]. (3) Data content requirements. These indicators mainly focus on the relevance, completeness, and accuracy of collected data. Relevance indicators include whether selected data are judged to be topic-related and whether there are relevance judgment criteria; completeness indicators include whether data have complete metadata descriptions and content completeness, including DRUM's sorting of data by time and relevance, Deep Blue Data's emphasis on metadata completeness [30], TDR's requirement that metadata descriptions meet standards and be complete [38]; UKDA emphasizes checking the accuracy of measurement data, using multiple measurements, observations, or sampling and expert verification to ensure data accuracy, and mentions the digitization degree of data and metadata [39]. (4) Data expression. ORNL DAAC emphasizes that data descriptions should be clear and easy to understand [37]; UKDA recommends using controlled vocabulary as much as possible during collection to reduce manual input [39]. (5) Data reuse. These indicators mainly concern the reusability and replicability of data utilization. Deep Blue Data recommends that data include descriptive metadata and should be reusable by others [30]; DataONE recommends that research results be replicable by others [40]; DRUM states that all data must be reviewed to ensure reusability, and data without reuse functionality may not be accepted by the repository [41].

3.3 Data Analysis and Processing Management

Data analysis and processing refers to the use of data processing software and hardware resources to process or analyze relevant data according to user needs and provide the resulting data products and analysis results to users in appropriate ways. The purpose of scientific data analysis and processing is to excavate and enhance the product value of scientific data, making it more discoverable (or findable), accessible, value-added, and interoperable [20]. Among the 15 selected data institutions, six mentioned indicators related to scientific data analysis and processing, with specific indicator content shown in Table 2 . After sorting, indicators in this stage can be summarized into four aspects: Data creation and description. This includes methods and standards for creating metadata, developing data dictionaries, and effective file naming. Data processing. This includes data processing depth and efficiency; data classification; whether data updates are timely; and whether visualization processing is possible. Data discoverability. This includes whether processing codes can be shared; whether data are easily retrievable; and whether file names are effective. Data usability. This includes whether processing codes are available for others to use; and whether datasets have value-added features.

3.4 Data Storage Management

The long-term preservation of scientific data requires repositories to have high security, with different scientific datasets having different content regarding security indicators. Data storage has technical requirements for repository systems and also requirements for storage content itself, including the confidentiality, integrity, and availability of storage formats and data content. Indicators from the investigated data institutions in this aspect can be divided into five categories: Data storage security. Deep Blue Data mentions that storage facilities should have appropriate disaster recovery functions and provide bit-level protection; it emphasizes the integrity and security during data transfer [30]; EROS mentions system security requirements and proposes that responsibility for IT security and privacy should be clearly assigned, and emphasizes that the importance of security protocols cannot be ignored [46]; TDR proposes that scientific data security must consider data backup, regular inspection, and providing resource service keys [47]. Data confidentiality. NCL Data mentions that file encryption technology should be used when storing data [48]. Storage operability. Operability emphasizes user-system interaction, reflected in whether users can obtain help when encountering difficulties. For example, NCL Data mentions that storage problem resolution requests can be submitted. Storage content indicators. Investigation reveals that these indicators mainly focus on whether data content remains accessible; whether storage content and systems are updated timely; whether data still have utilization value; whether data content can be preserved long-term; regulations on data storage volume; and providing different levels of data storage services. Storage format. Requirements for data storage formats, including whether formats are portable and diverse. For example, EROS believes that data storage formats should be portable and usable many years later [49].

3.5 Data Sharing and Utilization Management

The main drivers for promoting scientific data sharing abroad include: advancing scientific research; avoiding resource waste from duplicate research; effectively preserving scientific data long-term; and promoting scientific research collaboration and increasing citation rates and impact of research results [50]. China has not yet formed an effective data opening mechanism, and data sharing among government departments and scientific research institutions still faces barriers, creating "data silos" [51]. Investigation reveals that indicators involved by institutions in this stage mainly include six aspects: data publication, data citation, data openness, data impact, data usage legality, and data sharing privacy issues. Data publication. EROS emphasizes having publication format requirements; published products should include specified requirements, i.e., have completeness requirements; before publication, data accuracy, consistency, and completeness should be reviewed [52]. Data citation. This aspect mainly focuses on DOI issues, such as whether there are unified data citation standards or norms; whether DOIs are assigned to data; whether data citation formats are complied with; and whether guidelines or help for citing data are provided. Data openness. ICSU-WDS believes that data, metadata, products, and information for public domain use should be fully shared openly according to relevant laws and regulations [53]; both ARM and DRUM support free and open access. Data impact. NOAA emphasizes the scope of data audience and timeliness of dissemination; it also assesses whether data have substantial impact on important public sector or enterprise decision-making [42]. Data usage legality and data sharing privacy. TDR emphasizes that data users must not infringe on others' rights; must respect others' privacy; and must comply with all applicable local, state, national, and international laws and Texas Digital Library usage agreements [54]; ICSU-WDS emphasizes that data should comply with international ethical standards for research conduct; should comply with national or international laws and policies; data sharing should ensure certain privacy; and sensitive or restricted information should be marked when appropriate [53].

4. Construction of Scientific Data Quality Evaluation Model

From the perspective of various stages in scientific data management lifecycle, the author analyzed and summarized investigation results. Combining the requirements that evaluation index systems should have—systematicness, scientificity, conciseness, universality, and operability—a scientific data quality evaluation model was constructed from five dimensions: data management plan formulation, data collection management, data analysis and processing management, data storage management, and data sharing and utilization management (see Table 3). The indicators in this paper are abstracted and summarized based on the practices of the 15 investigated institutions, without exceeding these institutions' practices (relevant institutions are noted in Table 3). This index system has three hierarchical levels, and the constructed indicators follow a general-to-specific logical relationship, striving to fully reflect the characteristics of each stage of scientific data management lifecycle. Meanwhile, during the investigation, the author found that three indicators—usability, integrity, and objectivity—are common indicators running through all stages of scientific data management lifecycle. NOAA's explanation of these three indicators is relatively typical (see Table 4). When evaluating scientific data quality, not only should the specific content of individual indicators at each lifecycle stage be considered, but these three aspects should also be integrated. Therefore, during model construction, these three indicators were also incorporated to formulate specific evaluation content for each indicator.

Constructing a scientific data quality evaluation index system is important work for scientific data evaluation and management. The index system constructed in this paper considers the characteristics and objectives of each stage of the data lifecycle, hoping to provide reference and supplementation for scientific data institutions and platforms to establish scientific data evaluation systems. Because this index system involves all stages of scientific data lifecycle management, in practice, scientific data institutions can specifically draw on corresponding indicators according to the types, characteristics, and data management requirements of the data under their jurisdiction to construct evaluation indicators suitable for their institutions. The index system constructed in this paper still has issues such as insufficient experimental data and unsystematic analysis. The next stage of research will fully demonstrate the scientificity of this index system, use expert survey methods to design expert questionnaires, calculate weights for indicators at all levels through analysis, adjust and optimize the evaluation model, and provide reference for establishing decision-oriented scientific data center evaluation systems for national and local scientific data centers in China.

References

[1] NOAA. NOAA information quality guidelines [EB/OL]. [2019-10-27]. https://www.cio.noaa.gov/services_{programs}/info_{quality}.html.
[2] DANS. Evaluation of DANS EASY repository based on the FAIR Principles [EB/OL]. [2019-10-27]. https://dans.knaw.nl/en/about/organisation-and-policy/policy-and-strategy/Evaluation-of-DANS-EASY-based-on-the-FAIR-principles.pdf.

[3] WILKINSON M D, DUMONTIER M, AALBERSBERG I J, et al. Comment: the FAIR guiding principles for scientific data management and stewardship [J]. Scientific data, 2016, 3: 1-9.
[4] HU Cong. Research on the current situation, problems and countermeasures of scientific data submission management in China [J]. Science and Technology Entrepreneurship Monthly, 2019, 32(7): 81-84.
[5] STVILIA B, GASSER L, TWIDALE M B, et al. A framework for information quality assessment [J]. Journal of the American Society for Information Science and Technology, 2007, 58(12): 1720-1733.
[6] BATINI C, CAPPIELLO C, FRANCALANCI C, et al. Methodologies for data quality assessment and improvement [J]. ACM computing surveys, 2009, 41(3): 1-52.
[7] ZAVERI A, RULA A, MAURINO A. Quality assessment for linked data: a survey [J]. Semantic Web, 2016, 7(1): 63-93.
[8] KAHN M G, RAEBEL M A, GLANZ J M, et al. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research [J]. Medical care, 2012, 50(7): S21-S29.
[9] CHEN H, HAILEY D, WANG N, et al. A review of data quality assessment methods for public health information systems [J]. International journal of environmental research and public health, 2014, 11(5): 5170-5207.
[10] HUANG H, STVILIA B, JOERGENSEN C, et al. Prioritization of data quality dimensions and skills requirements in genome annotation work [J]. Journal of the American Society for Information Science and Technology, 2012, 63(1): 195-207.
[11] SHAO Yanhong. Research on the construction of evaluation index system for government open data quality in China [D]. Baoding: Hebei University, 2019.
[12] LI Xiaotong, ZHAI Jun, ZHENG Guifu. Research on data quality evaluation of local government open data in China—taking Beijing, Guangzhou and Harbin as examples [J]. Journal of Intelligence, 2018, 37(6): 141-145.
[13] LIU Guifeng, ZHANG Yu, LIU Qiong. Construction and case study of evaluation index system for scientific research data open platform [J]. Library and Information Knowledge, 2019(1): 21-31.
[14] ZHOU Yu, LIAO Siqin, RUAN Liping, et al. Research on construction and measurement of evaluation index system for data curation platform [J]. Research on Library Science, 2017(1): 35-42.
[15] YU Fangdong. International experience and reference for non-traditional data quality assessment [J]. Statistical Research, 2017, 34(12): 15-23.
[16] YU Houqiang, CAO Xueting. Research on the construction of altmetric data quality evaluation system [J]. Library and Information Knowledge, 2019(2): 19-27, 50.
[17] MA Feicheng, WANG Juncheng. Review of information lifecycle research (I) [J]. Journal of the China Society for Scientific and Technical Information, 2010(5): 939-947.
[18] DING Ning, MA Haoqin. Comparative study and reference of foreign university scientific data lifecycle management models [J]. Library and Infor-

mation Service, 2013, 57(6): 18-22.
[19] DATAONE. Data lifecycle [EB/OL]. [2020-02-09]. https://www.dataone.org/data-life-cycle.
[20] ZHANG Yang, XIAO Yanzhu. Interpretation and enlightenment of "Scientific Data Management Measures" from the perspective of lifecycle [J]. Research on Library Science, 2019(15): 37-43, 13.
[21] General Office of the State Council. Notice of the General Office of the State Council on issuing the Scientific Data Management Measures [EB/OL]. [2019-10-26]. http://www.gov.cn/zhengce/content/2018-04/02/content_{5279272}.htm.
[22] ORNL DAAC. Data management [EB/OL]. [2019-10-25]. https://daac.ornl.gov/data-management/.
[23] UCLA LIBRARY. Documentation and metadata overview [EB/OL]. [2019-10-26]. http://guides.library.ucla.edu/c.php?g=180580&p=1186345.
[24] ARM. Data management plan requirements [EB/OL]. [2019-10-26]. https://www.arm.gov/policies/datapolicies/digital-statement.
[25] EROS. Data management plans [EB/OL]. [2019-10-25]. https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/DMStrategyTemplate
[26] USGS. Data management: stewardship [EB/OL]. [2019-10-26]. https://www.usgs.gov/products/data-and-tools/data-management/stewardship.
[27] UCLA LIBRARY. About the DMP tool [EB/OL]. [2019-10-26]. http://guides.library.ucla.edu/c.php?g=180580&p=1190077.
[28] NEWCASTLE UNIVERSITY. Research data management: planning [EB/OL]. [2019-10-26]. https://research.ncl.ac.uk/rdm/planning/dmp/writing-a-data-management-plan/.
[29] WANG Dandan. Analysis of evaluation scale for scientific data management plan [J]. Library and Information Service, 2017, 61(18): 35-41.
[30] DEEP BLUE DATA. Policy and terms of use [EB/OL]. [2019-10-26]. https://deepblue.lib.umich.edu/data/agreement#preservation.
[31] UNIPROT. Guidelines for submitting updates or corrections to UniProtKB data [EB/OL]. [2019-10-26]. https://www.uniprot.org/help/submissions.
[32] DIGITAL CONSERVANCY. Policies and guidelines [EB/OL]. [2019-10-26]. https://conservancy.umn.edu/pages/drum/.
[33] TEXAS DATA REPOSITORY. Digital preservation policy [EB/OL]. [2019-10-25]. https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/291635428/Digital+Preservation-
[34] National Genomics Data Center. Big standards for big omics data [EB/OL]. [2019-10-26]. https://bigd.big.ac.cn/standards.
[35] CNGBdb. Submit data [EB/OL]. [2019-10-26]. https://db.cngb.org/cnsa/faq/#.
[36] CNGBdb. Review data [EB/OL]. [2019-10-26]. https://db.cngb.org/cnsa/faq/#.
[37] ORNL DAAC. Submit data [EB/OL]. [2019-10-26]. https://daac.ornl.gov/submit/#scope_{{{and}}{{acc
[38] TEXAS DATA REPOSITORY. Metadata dictionary [EB/OL]. [2019-10-25]. https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/493551668/Metadata+Dictionary.
[39] UK DATA SERVICE. Quality assurance [EB/OL]. [2019-10-26]. https://www.ukdataservice.ac.uk/manage-data/format/quality.aspx.
[40] DATAONE. Use data [EB/OL]. [2019-10-26]. https://www.dataone.org/use-data.

[41] DIGITAL CONSERVANCY. About the data repository [EB/OL]. [2019-10-26]. https://conservancy.umn.edu/pages/drum/.
[42] NOAA. NOAA information quality guidelines [EB/OL]. [2019-10-25]. https://www.cio.noaa.gov/services_{programs}/IQ_{{{Guidelines}}*{{103014}}}.html.*
[43] USGS. Data management: process and analyze—closely related activities [EB/OL]. [2019-10-25]. https://www.usgs.gov/products/data-and-tools/data-management/process-and-analyze-closely-related-activities?qt-science{{{support}}*{{page}}}*{{{related}}*{{con}}}=0#qt-science*{{{support}}*{{page}}}*{{{related}}*{{con}*
[44] TEXAS DATA REPOSITORY. Accessing and evaluating data [EB/OL]. [2019-10-25]. https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/289243266/Accessing+and+
[45] ARM. Data documentation [EB/OL]. [2019-10-26]. https://www.arm.gov/policies/datapolicies/data-documentation.
[46] USGS. Data management: backup & secure [EB/OL]. [2019-10-26]. https://www.usgs.gov/products/data-and-tools/data-management/backup-secure#tools.
[47] TEXAS DATA REPOSITORY. Information security [EB/OL]. [2019-10-25]. https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/292159828/Information+Security.
[48] NEWCASTLE UNIVERSITY. Research data management: working [EB/OL]. [2019-10-26]. https://research.ncl.ac.uk/rdm/working/.
[49] USGS. Data management: archiving [EB/OL]. [2019-10-26]. https://www.usgs.gov/products/data-and-tools/data-management/archiving?qt-science{{{support}}*{{page}}}*{{{related}}*{{con}}}=0#qt-science*{{{support}}*{{page}}}*{{{related}}_{{con}}}.
[50] HUANG Ruhua, QIU Chunyan. Review of foreign scientific data sharing research [J]. Information and Documentation Services, 2013(4): 24-30.
[51] XING Wenming, HONG Cheng. Openness as the norm, non-openness as the exception—interpretation of scientific data sharing and utilization in the "Scientific Data Management Measures" [J]. Library Tribune, 2019(1): 1-8.
[52] USGS. Data management: data release [EB/OL]. [2019-10-26]. https://www.usgs.gov/products/data-and-tools/data-management/data-release#elements.
[53] ICSU WORLD DATA SYSTEM. Data sharing principles [EB/OL]. [2019-10-26]. http://www.icsu-wds.org/services/data-sharing-principles.
[54] TEXAS DATA REPOSITORY. Terms of use [EB/OL]. [2019-10-25]. https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/289079299/Terms+of+Use.

Author Contributions

Jiang Hong: Determined the research topic and provided revision suggestions. Wang Chunxiao: Conceived the paper framework and wrote and revised the paper.

Academic Integrity Statement for Authors of *Library and Information Service*

Library and Information Service has always upheld the mission of publishing excellent academic paper results and promoting industry academic exchanges, and is committed to purifying the academic publishing environment and creating a good academic ecology. In 2013, the journal took the lead in formulating, publishing, and implementing the "Joint Statement on Library and Information Science Journals Abiding by Academic Ethics and Purifying the Academic Environment" (hereinafter referred to as the "Statement") (see: http://www.lis.ac.cn/CN/column/item202.shtml), and subsequently took the lead in formulating and publishing the "Joint Action Plan for Library and Information Science Journals to Resist Academic Misconduct" (hereinafter referred to as the "Joint Action Plan") (see: http://www.lis.ac.cn/CN/column/item247.shtml). To implement and realize this concept, this journal hereby solemnly declares that from now on, all contributing authors must promise that papers submitted to this journal must comply with the above "Statement" and "Joint Action Plan," consciously uphold academic ethics, and resolutely resist academic misconduct. *Library and Information Service* has zero tolerance for all papers suspected of plagiarism, piracy, and other forms of academic misconduct, and will adopt corresponding punitive measures.

Library and Information Service Editorial Office

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

Source: ChinaXiv — Machine translation. Verify with original.