A Maturity Assessment Model for Scientific Data Citation Standardization

Authors: Shi Yali, Zhao Tong, Wang Ting, Wang Ting

Date: 2023-05-23T00:00:00+00:00

Abstract

Purpose/Significance: To construct a maturity evaluation model for scientific data citation standardization and accelerate the normalization process of scientific data citation standards. This research facilitates the effective dissemination and sharing of scientific data achievements, and promotes the sustained and healthy development of scientific data citation standardization practices.

Methods/Process: Through literature review and integration of existing maturity theoretical models, maturity evaluation indicators for scientific data citation standardization practices are proposed; the Delphi method is employed to collect expert judgments on indicator importance; the Analytic Hierarchy Process is utilized to calculate indicator weights and derive a quantitative maturity calculation formula. Based on five levels—Initial Level, Process Management Level, Standardized Management Level, Quantitative Management Level, and Dynamic Optimization Level—a maturity evaluation model for scientific data citation standardization is constructed, and an applied analysis of the model is conducted using the Wuhan Public Data Open Platform as a case study.

Results/Conclusion: Through quantitative analysis, the data citation standardization practice of this data open platform is determined to be at the Standardized Definition Level; based on this finding, the survey results are analyzed and recommendations are proposed, providing reference and guidance for data managers, decision-makers, and users.

Full Text

Preamble

Research on the Maturity Evaluation Model for Scientific Data Citation Standardization

Shi Yali, Zhao Tong, Wang Ting
School of History and Culture, Hubei University, Wuhan 430062

Abstract:
[Purpose/Significance] This study constructs a maturity evaluation model for scientific data citation standardization to accelerate the normalization of scientific data citation standards. The research facilitates effective dissemination and sharing of scientific data outputs while promoting the sustainable and healthy development of citation standardization practices. [Method/Process] Through literature review and integration of existing maturity theoretical models, we propose maturity element indicators for evaluating scientific data citation standardization practices. The Delphi method was employed to collect expert judgments on indicator importance, and the Analytic Hierarchy Process (AHP) was used to calculate indicator weights and derive a quantitative maturity calculation formula. Based on five maturity levels—Initial, Process Management, Standardized Definition, Quantitative Management, and Dynamic Optimization—we construct a maturity evaluation model for scientific data citation standardization and apply it to the Wuhan Public Data Open Platform as a case study. [Results/Conclusions] Quantitative analysis reveals that the data citation standardization practice of this platform is at the Standardized Definition level. Based on these findings, we analyze the results and propose recommendations to provide reference for data managers, policymakers, and users.

Keywords: Scientific Data; Citation Standardization; Maturity Evaluation; Model
Classification Number: G252

With the advent of the big data era and continuous development of information technology, the frequency of data exchange and dissemination is accelerating. Against the backdrop of open science, scientific data citation standardization enables more regulated and rational data usage. Standardized citation of scientific data is of great significance, enhancing researchers' innovative vitality, creating greater academic value, and driving social development and progress. Current societal demands for comprehensive reform require scientific data citation standardization to keep pace with the times, making open and shared data more accessible to the public. In March 2023, the 8th Scientific Data Summit, themed "Scientific Data and New Paradigms of Research," explored new models, experiences, and pathways for innovation driven by scientific data. Notably, Percent Technology jointly released the first data capability advancement white paper with global authority IDC—the "Data Science Infrastructure Platform White Paper"—which facilitates efficient transformation of data into business knowledge to support decision-making and action, ultimately unlocking data value. China's National Standardization Committee issued "Information Technology—Scientific Data Citation" (GB/T 35294-2017), which stipulates that data users should follow standardized citation practices, adopt unified citation formats, methods, and metadata descriptions, and requires relevant data institutions to issue citation statements and principles to support standard implementation [1]. The General Office of the State Council's "Administrative Measures for Scientific Data" mandates that scientific data users comply with intellectual property regulations and properly cite scientific data in publications, patents, and monographs [2].

Establishing a maturity evaluation indicator system can scientifically and accurately measure the standardization level of scientific data citation practices, holding practical significance for both data platform development and regulating user citation behavior. Although numerous institutions and departments have developed their own scientific data citation standards—including citation principles, statements, and formats—varying enforcement intensity means existing guidelines only serve auxiliary functions, helping users understand and utilize data during platform access. Consequently, substantial room remains for advancing the standardization of scientific data citation. Based on this context, our study examines the current state of scientific data citation standardization domestically and internationally, grounds itself in maturity theoretical models, surveys expert opinions to establish a scientific evaluation indicator system, and clarifies assessment content, methods, and pathways. Using this indicator system, we investigate the Wuhan Scientific Data Open Platform, conduct statistical analysis of findings, and evaluate its citation standardization status.

This research is supported by the Ministry of Education Humanities and Social Sciences Youth Fund Project "Research on the Implementation Path of Scientific Data Citation Standardization in China" (Project No.: 20YJC870007) and the Hubei Provincial Department of Education Humanities and Social Sciences General Project "Research on the Maturity of Scientific Data Citation Standardization Driven by Data" (Project No.: 202011001301005).

Author Contributions: Shi Yali—conceptualization, data collection and analysis, manuscript revision; Zhao Tong—data collection and analysis, initial draft writing; Wang Ting—framework adjustment, content revision.

2.1 Major Achievements of Maturity Theoretical Models in Data Science Domains at Home and Abroad

To enhance software project capabilities, the Software Engineering Institute of Carnegie Mellon University developed the Capability Maturity Model (CMM) in 1988. In this model, management capability is divided into five levels, with the highest being the Optimizing level and the lowest the Initial level, sequentially including Managed (originally Repeatable), Defined, and Quantitatively Managed levels between them. Research demonstrates that this system effectively evaluates and improves software development capabilities [3]. With interdisciplinary integration, the data research field has gradually adopted maturity models as a new theoretical framework to enhance and evaluate data governance capabilities.

Currently, numerous mature theoretical maturity models exist in the data science field. Internationally, these include Syracuse University's Data Management Maturity Model (DM3) [4], the Enterprise Data Management Council (EDM) and Carnegie Mellon University's collaborative Data Management Maturity Model (DMM) [5], IBM's Data Governance Maturity Model [6], NASCIO's

Data Governance Maturity Model [6], Stanford University's Data Governance Maturity Model [7], the Australian National Data Service's (ANDS) Research Data Management Capability Maturity Model [8], and EDM Council's Data Management Capability Assessment Model (DCAM) [9]. Domestically, China has established its first national standard in data management, the "Data Management Capability Maturity Assessment Model" (GB/T 36073-2018, DCMM) [10], and the "Information Technology Service—Data Center Service Capability Maturity Model" (GB/T 33136-2016) [11].

Building upon existing models, many domestic scholars have proposed their own perspectives. For instance, Li Kepeng designed a big data security capability maturity model based on three dimensions (organization and personnel, technology and tools, processes and operations) and six stages of the data lifecycle [12]. Zhang Yaming et al. created a digital city governance maturity indicator system using factor and cluster analysis, surveyed data from 31 Chinese digital cities, and identified a four-gradient hierarchical model [13]. Qin Zhongyun integrated CMM levels, data governance capabilities, and the data asset lifecycle to determine assessment methods and planning routes for university library data governance maturity models [14]. Xu Lixin and Yuan Li established a "five-level, three-dimensional" maturity model for local government data open portals, determined measurement standards for each dimension, and completed a practical maturity assessment by evaluating eight provincial portals [15]. Ye Lan comparatively analyzed seven existing data management capability maturity models from five perspectives to derive comprehensive evaluation insights [16]. Cheng Fang et al. constructed a structural model comprising three dimensions—data quality capability, data quality level, and data quality elements—using a quality analysis method from application requirements to technical support [17].

2.2 Comparative Analysis of Maturity Theoretical Models in Data Science Domains at Home and Abroad

Comparative analysis of these maturity models reveals their respective applicability scopes, advantages, and disadvantages, providing valuable references for constructing a maturity evaluation indicator system for scientific data citation standardization practices. By summarizing models with comprehensive elements, complete evaluation criteria, clear processes, and high openness, we can build a new maturity evaluation indicator system. Our analysis shows that each maturity level in theoretical models has strict classification standards. Based on the CMM philosophy:

Maturity Level Classification: Most models adopt a five-level structure, except DCAM, which uses six levels by adding Level 0 (Unconscious) before the first level. In terms of applicability, DMM covers the broadest range, including IT, government, finance, aviation, and other sectors. IBM and DCAM primarily target financial and IT enterprises. Stanford's and ANDS's models mainly apply to personnel, policies, and technologies responsible for data governance within institutions, including university libraries, archivists, and computing facilities.

Both national standards DCMM and the Data Center Service Capability Maturity Model can evaluate organizations and institutions through self-construction, supervision, or external auditing.

Component Elements: DMM, ANDS, and DCAM share similarities in certain element settings, constructing basic maturity model components through strategy/policy, organization/personnel, architecture/platform, and governance/service dimensions, covering relatively comprehensive domains. Other models feature more targeted element selections. DCMM uses four elements—organization, system, process, and technology—making it more suitable for domestic data development contexts and more practical for organizational evaluation.

Openness: National standards DCMM and the Data Center Service Capability Maturity Model offer free full documentation with detailed maturity level classifications and capability domain evaluation methods, demonstrating high openness. ANDS also provides full model access for research and discussion. Other models lack free online access and require membership for support, resulting in relatively low openness.

3.1.1 Scientific Principle

Scientific rigor is the foremost principle in all research, requiring scientific methods, approaches, and procedures. The maturity evaluation indicators for scientific data citation standardization practices are established on this foundation. During indicator extraction, we ensure scientific validity by referencing relevant literature and employing expert surveys, enabling the indicators to reflect and represent characteristics at each maturity stage, ultimately establishing the required element indicators for the evaluation system.

3.1.2 Applicability Principle

Indicator selection for evaluating scientific data citation standardization practices should be context-specific, with adjustments made to the constructed indicators based on full consideration of actual assessment objects to enhance practical applicability. Maturity evaluation system indicators possess interrelatedness and form an inseparable unity; therefore, quantitative factors should be fully considered during indicator selection and system construction.

3.1.3 Operability Principle

The ultimate purpose of constructing a maturity evaluation indicator system for scientific data citation standardization is to assess selected research objects, making operability an indispensable principle. Based on comprehensive research process considerations, maturity element indicators should be widely distributed and easily collected, while maturity level classifications should be concise, clear, and strongly characteristic, ensuring stronger operability in subsequent surveys and calculations.

3.2 Establishment of the Scientific Data Citation Standardization Maturity Evaluation Indicator System

Our extraction of maturity element indicators primarily involved reviewing existing maturity models and literature on data citation. By refining process domains or functional domains from maturity theories, we initially formed element indicators rooted in existing research. First, based on the eight domestic and international maturity models mentioned above, we referenced, borrowed, and integrated them to construct four first-level indicators: institutional guarantee, citation management, relevant stakeholders, and technical support. Second, to achieve internal coherence within first-level indicators while demonstrating interconnections between components, we reviewed literature on scientific data citation, selecting representative references based on citation frequency and relevance as samples for indicator extraction. We summarized data citation-related content from these documents to identify 17 second-level indicators. Finally, we classified these indicators into the aforementioned first-level categories based on their relationships and distinctions, completing the preliminary selection of maturity element indicators (see Table 1).

3.3.1 Questionnaire Design

The Delphi method was employed to determine maturity element indicators, with measurement conducted using a Likert scale questionnaire consisting of two parts. The first part collected basic information about respondents, including professional identity, background, education level, and perspectives on scientific data citation standardization practices. The second part explained the scale design, using a 1-9 scoring system for constructing judgment matrices. Based on this standard, experts judged the importance of first-level and second-level indicators through pairwise comparisons, with scoring results representing relative importance between indicator pairs to reduce uncertainty from subjective judgments. Selected experts represented diversified research directions and work experiences to ensure scientifically valid results for analysis.

3.3.2 Questionnaire Distribution and Data Collection

This study employed the Delphi method, distributing questionnaires to experts in data citation-related fields to collect their judgments and evaluation suggestions on element indicators. Survey results were organized and analyzed to finalize the evaluation indicator system for scientific data citation standardization practices.

Data collection combined offline and online methods, with 56 questionnaires distributed, 25 returned, and 19 valid responses after eliminating invalid questionnaires. Among respondents, 4 held bachelor's degrees, 9 master's degrees, and 6 doctoral degrees (see Table 2). Selected experts possessed research achievements or practical experience in scientific data citation and recognized the need for more standardized citation practices. Bachelor's degree holders had extensive industry experience in data citation, making their responses valid for statistical purposes. For clearer subsequent analysis, valid questionnaires were numbered as Pn, where n represents the chronological order of receipt.

3.4 Determination of Evaluation Indicator Weights

The Analytic Hierarchy Process (AHP) was used to determine indicator weights, primarily quantifying expert scores to derive numerical values and verifying allocation rationality through consistency checks. AHP consists of four steps: establishing the hierarchical structure model, constructing judgment matrices, calculating hierarchical single sorting with consistency testing, and computing hierarchical total sorting to obtain comprehensive weights.

3.4.1 Establishing the Hierarchical Structure

In AHP, decision objectives, influencing factors, and decision objects constitute the top, middle, and bottom layers, respectively. The top layer (objective layer) is "Maturity of Scientific Data Citation Standardization Practice," designated as A. The middle layer (criteria layer) comprises four first-level indicators: institutional guarantee, citation management, relevant stakeholders, and technical support, designated as Bn. The bottom layer (measures layer) includes 17 operational second-level indicators for achieving the top-level objective, designated as Cn (see Figure 1 [Figure 1: see original paper]).

3.4.2 Constructing Comparison Judgment Matrices

After establishing the hierarchical structure, expert scores on indicator importance were compiled to construct judgment matrices using relative scales for pairwise comparisons, reducing difficulties and uncertainties from simultaneous scoring of all indicators. For elements under a specific objective or criterion, pairwise comparisons follow importance evaluation scales, where aij represents the comparison result between element i and element j, ensuring aij = 1/aji and aii = 1. Table 3 presents nine importance levels and their assignments for matrix element aij.

Based on 19 experts' pairwise comparison scores, the following mean judgment matrices were constructed:

(1) **First-level indicator judgment matrix:** Pairwise relative importance comparisons of criteria layer indicators to the objective layer for the four first-level indicators: institutional guarantee, citation management, relevant stakeholders, and technical support (see Table 4).

(2) **Second-level indicator judgment matrices:** Pairwise relative importance comparisons of measures layer indicators to criteria layer. For institutional guarantee's three second-level indicators (policies and regulations, citation norms, relevance explanation), see Table 5 . For citation management's four indicators (citation object type, storage format, data quality and readiness, citation elements), see Table 6 . For relevant stakeholders' seven indicators (research and education institutions, data storage institutions, government functional departments, database vendors and commercial operators, other non-profit organizations, end-user feedback), see Table 7 . For technical support's three indicators (platform construction, citation software, identity authentication and recognition), see Table 8 .

3.4.3 Hierarchical Total Sorting and Consistency Testing

Based on the mean judgment matrices from survey results, importance calculations can use methods including the sum-product method, root method, power method, and least squares method. This study employs the root method for weight vector calculation. For consistency testing, CI is compared with the average random consistency index RI to derive the test statistic CR. Since AHP is a subjective weighting method, expert scoring may contain errors, necessitating consistency verification. Using MATLAB software, hierarchical single sorting and consistency testing were performed on first-level indicators, followed by hierarchical total sorting and consistency testing.

(1) **Calculate row product Mi for judgment matrix:**

$$\begin{aligned}
\text{Institutional Guarantee}(M_1) &= 1 \times 0.3971 \times 0.5556 \times 1.2038 = 0.2656 \\
\text{Citation Management}(M_2) &= 2.5185 \times 1 \times 0.8571 \times 1.0963 = 2.3665 \\
\text{Relevant Stakeholders}(M_3) &= 1.8000 \times 1.1667 \times 1 \times 0.5921 = 1.2434 \\
\text{Technical Support}(M_4) &= 0.8307 \times 0.9122 \times 1.6889 \times 1 = 1.2798
\end{aligned}$$

(2) **Calculate nth root Wi of Mi:**

$$\begin{aligned}
\text{Institutional Guarantee}(k_1) &= \sqrt[4]{0.2656} = 0.7179 \\
\text{Citation Management}(k_2) &= \sqrt[4]{2.3665} = 1.2403 \\
\text{Relevant Stakeholders}(k_3) &= \sqrt[4]{1.2434} = 1.0560 \\
\text{Technical Support}(k_4) &= \sqrt[4]{1.2798} = 1.0636
\end{aligned}$$

(3) **Normalize vector W = (Wk1 Wk2 Wk3 Wk4):**
The judgment matrix eigenvector is obtained.

(4) **Consistency test:** The consistency test result shows CR < 0.10, indicating satisfactory consistency in the expert judgment matrix and acceptable

weight vectors. Similarly, B1-C, B2-C, B3-C, and B4-C random consistency ratios are all below 0.10, passing consistency tests and yielding acceptable weights.

Comprehensive weights are calculated by multiplying first-level and second-level indicator weights (see Table 9). Combined with survey-based indicator scores, the maturity calculation formula is determined.

3.5 Maturity Level Classification for Scientific Data Citation Standardization

Maturity level classification reflects the progressive state evolution of scientific data citation standardization from low to high levels. Based on current research and development of maturity models, combined with the status quo of scientific data citation standardization practices and referencing data management capability maturity level classifications, this study divides maturity into five sequential levels: Initial, Process Management, Standardized Definition, Quantitative Management, and Dynamic Optimization (see Figure 2 [Figure 2: see original paper]).

Initial Level: The starting point where user data citation is spontaneous individual behavior based entirely on personal needs. Citation platforms remain unstable, lack unified management processes, and related infrastructure is under construction. As citation awareness grows and platforms develop, maturity advances to the second level.

Process Management Level: The importance and significance of data citation gain broader attention, transforming from individual to collective behavior. Data management begins establishing relevant regulations, citation procedures become gradually process-oriented, and citation technologies improve but remain incomplete. With strengthened supervision mechanisms and enhanced management service awareness and capabilities, maturity advances to the third level.

Standardized Definition Level: Scientific data citation processes become institutionalized, with scientifically effective citation standards and rules established, gradually forming a standard system. Platforms emphasize relationships between service providers and users during citation, highlighting data awareness and practical significance. As citation standard systems continuously improve, maturity advances to the fourth level.

Quantitative Management Level: Scientific data citation and management capabilities reach a high level. During data sharing and dissemination, citation follows formal variable management processes, and scientific data is gradually incorporated into research performance evaluation systems. Service providers and users engage in bidirectional management and supervision, refining citation standardization systems. Through further optimization, maturity advances to the final level.

Dynamic Optimization Level: Scientific data citation becomes normalized. This continuous improvement stage involves constant refinement through practice and feedback, achieving real-time workflow optimization through multifaceted supervision to enhance citation quality. This stage maintains optimal citation processes and regulatory mechanisms to support further citation activities.

4.1 Survey Purpose and Object

Based on the constructed evaluation model, we selected an appropriate data open platform for empirical research to test the indicator system's scientific validity and operability. Through indicator scoring and data analysis, we obtained final maturity results to identify the platform's citation standardization status and propose effective improvement measures.

This study selected the Wuhan Public Data Open Platform as the survey object for the following reasons: (1) As a government-launched platform with a gov.cn domain, it enjoys official recognition at the prefecture-level administrative tier, offering data for public download or API access. (2) As of the survey, the platform had opened 60 municipal departments/districts, 1,406 open catalogs, 1,036 data APIs, 11,864 data items, and 295.7 million data records, comprising a massive collection covering various fields, themes, and industries with application, statistical, and interactive features that enable comprehensive information collection.

4.2 Maturity Evaluation of Wuhan Public Data Open Platform's Data Citation Standardization

The Wuhan Public Data Open Platform primarily provides data download, API development services, and application upload/download functions, covering tabular, textual, image, map, multimedia, and other real-time and non-real-time government public data [28]. The platform strengthens management in citation policy and data security to achieve application standardization. This study investigates the platform's data citation standardization status to provide a basis for subsequent indicator evaluation and maturity calculation.

4.2.1 Institutional Guarantee

Since September 2020, the platform has published draft data resource management measures and formally issued the "Public Data Resource Management Measures" in November 2021, which explicitly regulates data resource management activities within the municipality and distinguishes management approaches for public versus state-secret data. The draft solicited public opinions and suggestions from citizens, organizations, and institutions regarding data citation issues, with revisions made based on feedback to ensure rational data resource application. For citation norms, the platform issued notices and documents on API usage specifications and core metadata standards, though unified citation formats were not explicitly addressed. Relevance explanations cover potential service issues, including disclaimers and privacy protection statements that clearly define rights and obligations between the platform and users, such as: "All platform content is owned by Wuhan Municipal People's Government," "The platform may terminate services and notify users when legal violations are suspected," and "Users have the right to access all government data resources but may not transfer them for compensation."

4.2.2 Citation Management

The platform's data demonstrates diversity across multiple types and formats, including CSV, XLSX, JSON, XML, and RDF, enhancing machine readability and facilitating multi-format data access and processing. For large datasets that would increase electronic access barriers through form-only retrieval, the platform provides API access without plugin or browser limitations, enabling convenient data acquisition [29]. Datasets are categorized by field, theme, and industry according to national standards and relevant regulations, ensuring convenient user access and citation.

In April 2021, the platform piloted the "Wuhan Public Data Resource Open Core Metadata Standard," which specifies core metadata and representation methods required for describing open data resource catalogs, applicable to Wuhan government departments and public service enterprises' public data resource open management and application. Core metadata includes identifiers, names, abstracts, publication dates, file sizes, data provider information (codes, contact details, administrative regions), open attributes (type and conditions), and data item information (names and types).

4.2.3 Relevant Stakeholders

Multiple stakeholders collaborate throughout the data lifecycle during citation. Wuhan Public Data Open Platform's partners include two categories: 44 municipal departments and 16 district-level divisions, each aggregating corresponding datasets to enhance citation convenience.

The platform emphasizes user interaction, establishing dedicated taskbars to address user needs and improve satisfaction. After logging in, users can request unopened data by simply completing forms with requirement titles, content, data purpose, personal information, and expected format. The website features survey questionnaires on service usability and data needs, with dozens of participants likely reflecting regular visitors. Users can also submit inquiries and suggestions regarding data issues, access, platform experience, and complaint services. The platform lists feedback chronologically with responses, publicly discloses corrective issues including data updates, catalog problems, access permissions, and garbled text, and follows up on operations to resolve usage problems promptly. By enhancing management service awareness and enabling multi-channel expert collaboration, the platform meets diverse user data needs.

4.2.4 Technical Support

The platform established specialized departments for daily supervision and inspection, with accountability mechanisms to identify and improve platform construction issues promptly. The Wuhan Cyberspace Information Office published annual government website work reports for 2021-2022, providing detailed registration of platform security protection including security assessment frequency, problem identification and rectification quantities, emergency drill implementation, and security responsibility designation.

Following strict classification requirements covering security, personal/enterprise information protection, and application demands, the platform categorizes public data resources into different security levels using identity authentication and recognition technology to mitigate potential impacts or damages on individuals, organizations, societal interests, or government operations during citation. The platform adheres to classification principles—compliance, security, operability, timeliness, and extensibility—dividing open types into "non-open," "conditionally open," and "unconditionally open" with corresponding explanations.

4.3.1 Scoring Method

This study's scoring approach is based on the maturity evaluation indicator system levels and descriptions described above, establishing evaluation criteria for second-level indicators across different maturity levels. Since all stakeholder indicators except "end-user (feedback)" share the same type, identical evaluation standards were applied. Scientific data citation standardization must satisfy all requirements of previous levels before reaching the next (see Table 10) [30].

4.3.2 Maturity Calculation

Based on the maturity indicator level evaluation criteria, relevant data was collected from the platform. Two rounds of weighting calculations yielded specific scores for second-level indicators and the platform's final total score. Let the final score be S, with the four first-level indicators—institutional guarantee, citation management, relevant stakeholders, and technical support—scored as S1, S2, S3, and S4 respectively. S1 is calculated as: S1 = W1 × (R is indicator assignment), with S2, S3, and S4 calculated similarly. Using this formula, each indicator was assigned values to compute the Wuhan Public Data Open Platform's comprehensive evaluation score (see Table 11).

The Wuhan Public Data Open Platform's maturity value is: S = S1 + S2 + S3 + S4 = 3.0550. According to the scientific data citation standardization maturity model evaluation criteria, the platform's citation standardization maturity is currently at the **Standardized Definition level**.

4.4 Results Analysis and Recommendations

Calculations indicate that the Wuhan Public Data Open Platform's citation standardization practice is at the Standardized Definition level, with notable improvements in institutional guarantee, citation management, and technical support over the Process Management level, and some stakeholder development. The platform is trending toward Quantitative Management level, demonstrating strong overall citation standardization practice but with substantial room for future improvement.

Institutional Guarantee: The platform has relatively comprehensive policy and regulation development, soliciting public modification suggestions before formal document release, establishing a standard institutional framework, and revising policies through quantitative assessment to enhance applicability and rationality. Service agreements including disclaimers, rights and obligations, and privacy protection supplement policies where needed. However, the platform has not explicitly issued citation norm documents, potentially causing errors or non-standard practices regarding citation elements (author, title, institution, date, identifier). The platform should promote national standard GB/T 35294-2017 applications or establish dedicated data working groups to develop context-appropriate citation norms through expert research, project discussions, and empirical investigations, followed by public solicitation and revision based on feedback.

Citation Management: Citation object types and storage formats demonstrate diversity. Users can access related datasets or items when retrieving data and obtain full datasets via API without plugin or browser limitations, avoiding download barriers for large volumes. Data quality and readiness are adjusted based on regular monitoring and user feedback. However, the platform applies uniform citation management across different domains without distinguishing management by data characteristics or lifecycle phases, potentially creating significant differences in citation objectives and requirements for specialized data [31]. The platform should build a citation management system coordinating object types and formats, enhance professionalism and standardization, organize citation elements, adopt metadata standards for citation processes, and implement long-term metadata preservation.

Relevant Stakeholders: Various institutions and departments participate, with research and education institutions, data storage institutions, publishing organizations, and government functional departments showing higher engagement and clearer responsibilities. Database vendors and other commercial operators, along with non-profit organizations, participate less frequently. The platform emphasizes end-user feedback through dedicated interaction channels, multiple contact methods (system, phone, email), and public feedback lists and error corrections, bridging platform-user communication. The platform should strengthen multi-stakeholder collaboration, encourage all parties to participate in citation processes, monitor citation status in real-time, and form multi-stakeholder alliances. For feedback, the platform could add push notifications and monitoring services, promptly delivering solutions to users and customizing related descriptive service information, while tracking visit frequency and account status to statistically analyze citation patterns and design data citation update strategies [32].

Technical Support: The platform has established relatively complete accountability mechanisms with designated personnel, timely problem identification, and improvement. Public data resources are classified by security levels using identity authentication technology to ensure citation security. However, citation technology lags, with only some departments or users employing software for citation work. With citation norms not yet implemented, the platform could reference other data sharing platforms to automatically generate citation information for datasets through data citation resolution systems, enabling users to view citation information on retrieval pages for direct use. Additionally, the platform could provide data citation identifier resolution services, allowing users to input unique identifiers in search bars to access data pages or metadata description pages [33].

In summary, this study preliminarily established a scientific data citation standardization maturity evaluation model and applied it in practice. However, as data resources continuously update and user demands increase, maturity evaluation standards are not static; element selection and level classification require ongoing adjustment and updating. To address limitations, future improvements will focus on three aspects: First, refining the indicator system by selecting experts from diverse universities, enterprises, or data management organizations to evaluate indicator scientificity and establish more rational maturity indicators. Second, expanding research scope, as Wuhan's districts and counties will establish new data open platforms with advancing citation work, requiring enhanced empirical exploration to strengthen the indicator system's general applicability.

References

[1] China National Standardization Committee. Announcement No. 32 of 2017 National Standards [EB/OL]. [2020-01-30]. http://www.sac.gov.cn/gzfw/ggcx/gjbzgg/201732/.

[2] General Office of the State Council. Notice on Issuing Administrative Measures for Scientific Data [EB/OL]. [2018-04-02]. http://www.gov.cn/zhengce/content/2018-04/02/content_{5279272}.htm.

[3] CMMI Institute. Data Management Maturity Model [DB/OL]. [2019-03-18]. https://cmmiinstitute.com/store/data-management-maturity-(dmm).

[4] Ye Lan. Analysis of Research Data Management Capability Maturity Model [J]. Library and Information Service, 2015(02): 115-123.

[5] EDM Council. Data Management Maturity Model [EB/OL]. [2020-03-12]. https://max.book118.com/html/2018/1229/7161105013001200.shtm.

[6] NASCIO. NASCIO: Data Governance Part : Maturity Models-A Path to Progress [EB/OL]. [2014-12-04]. http://www.docin.com/p-978629039-f2.html USA.

[7] Stanford University Data Governance Maturity Model [EB/OL]. [2020-03-12]. http://www.doc88.com/p-7098899678378.html.

[8] ANDS. Research Data Management Framework: Capability Maturity Guide [EB/OL]. [2014-05-29]. http://www.ands.org.au/guides/dmframework/dmf-capability-maturity-guide.html.

[9] EDM Council. The Data Management Capability Assessment Model [EB/OL]. [2019-08-12]. http://www.fuduo.wang/index.php/2019/08/12/100.html.

[10] China National Standardization Administration. Data Management Capability Maturity Assessment Model: GB/T 36073-2018 [S]. Beijing: Standards Press of China, 2018: 1-39.

[11] China National Standardization Administration. Information Technology Service—Data Center Service Capability Maturity Model: GB/T 33136-2016 [S]. Beijing: Standards Press of China, 2016: 1-40.

[12] Li Kepeng, Mei Jingting, Zheng Bin, et al. Research on Big Data Security Capability Maturity Model Standards [J]. Information Technology and Standardization, 2016(07): 59-61.

[13] Zhang Yaming, Pei Lin, Liu Haiou. Empirical Research on Digital City Governance Maturity in China [J]. China Science and Technology Forum, 2010(05): 70-76.

[14] Qin Zhongyun. Research on University Library Data Governance and Maturity Model in Big Data Environment [J]. New Century Library, 2019(11): 62-67.

[15] Xu Lixin, Yuan Li. Maturity Assessment Research on Local Government Data Open Portals [J]. Library and Information Service, 2019, 63(12): 52-58.

[16] Ye Lan. Comparative Research and Implications of Data Management Capability Maturity Models [J]. Library and Information Service, 2020, 64(13): 51-57.

[17] Cheng Fang, Zhao Yanqing, Wang Lei. Research on Data Quality Capability Maturity Model Based on Data Service Platform [J]. Standard Science, 2020(10): 120-123.

[18] Si Li, Xing Wenming. Investigation of Foreign Scientific Data Management and Sharing Policies and Implications for China [J]. Information and Documentation Services, 2013(01): 61-66.

[19] Wang Siming, Ning Zichen. Analysis of Stakeholder Roles in Scientific Data Citation Process [J]. Digital Library Forum, 2019, (04): 29-36.

[20] Xiang Ying, Lai Jianfei, Ding Ning. Practice Exploration of Scientific Data Management Services in University Libraries—Taking Wuhan University Social Science Data Management as an Example [J]. Information Studies: Theory & Application, 2013, 36(12): 89-93.

[21] Xia Yikun. Discussion on Government Data Governance in Open Data Environment: Conceptual Framework and Main Issues [J]. Library and Information Service, 2018(01): 95-104.

[22] Zhang Peifeng, Zhang Lianfen. Innovation of Library Scientific Data Management Services Under Global Research Paradigm Transformation—Based on Data Management Lifecycle Perspective [J]. Library Theory and Practice, 2019(05): 39-48.

[23] Liu Zhifeng, Wang Jimin, Li Qian. Review of Metadata Quality Evaluation Research [J]. Information Studies: Theory & Application, 2022, (12): 1-16.

[24] Gu Liping. Data Management Services in Research Paradigm Transformation: Approaches to Achieving Open Access, Open Data, and Open Science [J]. Journal of Library Science in China, 2018, 44(06): 43-58.

[25] Shi Yali, Zhao Xueqin. Analysis of Relationships Among Scientific Data Citation Stakeholders from Cooperative Game Perspective [J]. Digital Library Forum, 2019, (01): 15-20.

[26] Zhao Jiajun. Research on Data Citation Recognition in Scientific Literature [D]. Nanjing Agricultural University, 2019.

[27] Wei Yue, Liu Guifeng. Analysis of Foreign University Scientific Data Management and Sharing Policies Based on Data Lifecycle [J]. Journal of Intelligence, 2017, 36(05): 153-158.

[28] Chen Yaping. Comparative Research on Chinese and Foreign Government Open Data Practices and Management Mechanisms [D]. Nanjing University, 2014.

[29] Hu Yifang, Lin Yan. Policy and Regulation Guarantees for Canadian Government Data Openness and Implications for China [J]. E-Government, 2017, (05): 2-10.

[30] Niu Lixue. Research on Government Data Governance Maturity Model [D]. Hebei University, 2020.

[31] You Yi. Review of Domestic Library Data Services Research in the Past Decade [J]. Library Research and Work, 2022, (03): 55-60.

[32] Zhong Yan, Zhong Yanling, Guo Wensi. Research on Scientific Data Sharing Construction [J]. Application of Electronic Technique, 2021, 47(09): 64-68+74.

[33] Zhu Yanhua, Hu Lianglin, Kong Lihua, et al. Development and Promotion of National Standards for Scientific Data Citation [J]. e-Science Technology & Application, 2018, 9(06): 25-30.

Author Contributions: Shi Yali—conceptualization, data collection and analysis, manuscript revision; Zhao Tong—data collection and analysis, initial draft writing; Wang Ting—framework adjustment, content revision.

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