

# 基于ISM-MICMAC的科学数据引用标准化 影响因素及作用机理研究\*

史雅莉 沈心悦 孙蕴哲  
(湖北大学历史文化学院, 武汉 430062)

**摘要:** 数据密集型范式的到来使得数据作为一项战略性资源, 在科研及各行各业中发挥越来越重要的作用。系统探究科学数据引用标准化的内外部影响因素, 对于推动科学数据引用标准化实践的发展显得尤为重要。运用ISM-MICMAC分析法, 剖析各影响因素的指标特性和不同影响因素间的作用机理, 构建科学数据引用标准化影响因素层级关系结构。基于此, 提出科学数据引用标准化的推行策略: 加强科学数据质量监管和治理, 为科学数据引用标准化奠定基础; 落实科学数据引用标准规范, 兼顾不同学科引文风格的差异性; 依托现有引文技术标准, 实现科学数据引文创建与管理的自动化; 明确相关利益主体间权责关系, 协同推进科学数据引用标准化实践发展。

**关键词:** 科学数据; 科学数据引用; 标准化; 影响因素

**中图分类号:** G252 **DOI:** 10.3772/j.issn.1673-2286.2024.05.009

**引文格式:** 史雅莉, 沈心悦, 孙蕴哲. 基于ISM-MICMAC的科学数据引用标准化影响因素及作用机理研究[J]. 数字图书馆论坛, 2024, 20(5): 79-88.

随着计算机通信技术和科研范式的转变, 知识传播的形式也在逐步发生变化。期刊论文和学术专著在知识传播中的主导地位开始受到海量开放数据及其集合的挑战。2018年3月, 国务院办公厅颁布了《科学数据管理办法》, 进一步加强和规范科学数据管理。2023年3月, 数据科学峰会在北京举行, 此次大会以“数据·进阶”为主题, 邀请权威专家学者解读数据科学、数字政府建设及数字化业务, 推出一站式数据价值实现平台, 并发布《用数据创造价值 以智能激发增长——数据科学基础平台白皮书》。2023年4月, 第八届科学数据大会暨CODATA中国全国委员会学术年会召开, 会议围绕“科学数据驱动科研新范式”“科学数据关键技术及应用”“科学数据赋能数字经济”“科学数据政策与基础设施”4个主题展开, 探究科学数据与科研新

范式, 强调科学数据资源再利用的重要性。

科学数据引用标准化是实现科学数据共享的重要途径之一, 同时也是认可和尊重数据创建者学术贡献, 避免学术不端行为的重要手段。在科学数据引用标准化的实践发展方面, 国外如德国科学基金会(Deutsche Forschungsgemeinschaft, DFG)、英国数字市场部(Digital Markets Unit, DMU)、哈佛大学社会科学研究院(Harvard Graduate School of Social Sciences)等机构陆续开展科学数据引用相关的项目实践, 并发布科学数据引用行为规范和要求。DataCite是一家于2009年在英国成立的国际非营利性组织, 由多家机构联合发起, 包括英国大英图书馆、丹麦技术信息中心、荷兰代尔夫特理工大学图书馆等。其为科学数据创建元数据集, 为用户分配DOI, 增强数据搜索能力, 与会

收稿日期: 2024-02-15

\*本研究得到教育部人文社会科学青年基金项目“我国科学数据引用标准化的实现路径研究”(编号: 20YJC870007)资助。

员共同建立共享数据库和高效引用机制。DataCite于2009年11月推出国际数据引用倡议，截至2022年已完成第3次修订。国内科学数据引用工作起步较晚，2017年12月，国家标准化委员会颁布了《中华人民共和国国家标准公告（2017年第32号）》，其中《信息技术 科学数据引用》（GB/T 35294—2017）列为被批准的国家标准之一，并于2018年7月正式实施。

综上所述，有关科学数据规定与国家标准的颁布以及学术会议的召开，推动了科学数据的发展与完善，科学数据引用标准化问题也逐渐受到关注。但是，我国科学数据引用在引用规范的实施、引用效果评价等方面仍面临诸多挑战<sup>[1]</sup>。系统分析科学数据引用标准化的影响因素，有助于提高科研工作者的数据素养，加深其科学数据引用意识，为相关科学数据管理机构提供决策支持，同时为科学数据引用标准化的方法与技术提供理论支持。

## 1 研究方法与研究过程

ISM-MICMAC分析法是一种结合使用解释结构模型（Interpretative Structural Model, ISM）和交叉影响矩阵相乘法（Matrixed Impacts Corises-Multiplication Appliance Classment, MICMAC）的方法。ISM由美国学者J. 华费尔特博士于1973年提出，它将一个复杂的系统分为几个子系统，其中有向图和关联矩阵是主要形式，通过使用计算机分析大量实践经验和知识，形成一个多层次的层次结构模型，可将模糊的想法、观点转化为一种直观的、存在结构关系的模型，为系统管理者提供决策依据<sup>[2]</sup>。该方法自提出后在工程管理、电子商务、数据科学等不同的学科领域得到了广泛应用，用于分析主题目标与影响因素的关联性。本研究应用这一方法分析科学数据引用标准化各影响因素间的关联性及其作用机理，研究方法与研究主题具有较高的契合度。同时在ISM分析的基础上，运用MICMAC进一步开展定量分析，通过坐标系表示出各因素之间的影响与依附关系。

具体研究过程如下：①通过德尔菲法，征询专家意见，识别科学数据引用标准化的影响因素；②利用MATLAB软件构建邻接矩阵和可达矩阵，确立各影响因素间的直接作用关系和间接作用关系；③对可达矩阵进行分解，明确科学数据引用标准化影响因素的层

级关系结构，生成图谱；④通过MICMAC分析，判断科学数据引用标准化影响因素的指标特性；⑤归纳研究结论，提出科学数据引用标准化的推行策略。

## 2 科学数据引用标准化影响因素的识别与指标体系构建

通过系统的文献调研，初步确定科学数据引用标准化的影响因素，以此为基础设计量表内容，进行专家调研，最终构建科学数据引用标准化影响因素指标体系。

### 2.1 科学数据引用标准化影响因素识别

（1）通过文献调研初步确定科学数据引用标准化的影响因素。文献调研阶段指标具体选取流程如下。①文献检索：设定检索式“Data Cite” OR “Data Citation” AND “Influencing Factors” OR “Data Reuse” OR “Dataset Reuse” AND “Influencing Factors”，以及“科学数据引用” OR “数据引证” AND “影响因素”。检索源选择Web of Science核心合集、Scopus文摘数据库、Proquest硕博学位论文全文库、中国知网数据库，在数据库中对科学数据引用相关主题文献及其引文进行追踪。②文献选取标准：文献类型包括中英文期刊论文、会议论文、学位论文；将同一作者的相似研究合并，保留学位论文，若期刊论文和会议论文是博士学位论文研究内容的延伸则保留。③内容分析与指标提炼：从纳入研究范围的文献中抽取影响因素指标内容、指标关联度描述等信息，并用Excel表格进行整理。④规范化处理：结合相关主题词表，对指标术语表达进行规范化处理，消除歧义。

（2）通过调查问卷对影响因素进行二次识别和关联判断。本次调查邀请来自不同学科领域的长期从事学术研究的专家学者填写问卷，对收集到的25份调查结果进行数据清洗和预处理，剔除内容缺失、专家信息补充失败等无效数据，通过理论饱和度检验，最终获得16份有效调查结果。

调查问卷第一部分涉及被调查者的基本情况，包含被调查者的性别、学位、专业背景、职业、工作年限、科学数据引用频率、对科学数据引用标准化工作的认知态度等信息，调查结果见表1。

表1 被调查者的基本情况

序号	性别	学位	专业背景	职业	工作年限	引用频率	认知态度
E1	女	硕士	管理学	教辅工作者	1~3年	一般	不确定
E2	女	硕士	管理学	教师	1年及以下	一般	支持
E3	女	博士	医学	临床工作者	1~3年	经常	非常支持
E4	女	博士	管理学	科研人员	20年及以上	经常	非常支持
E5	女	博士	物理学	教师	1~3年	一般	非常支持
E6	男	博士	管理学	教师	6~10年	经常	非常支持
E7	女	硕士	工学	企业职员	1年及以下	一般	不确定
E8	男	硕士	管理学	教辅工作者	6~10年	经常	非常支持
E9	女	博士	管理学	教师	1~3年	一般	非常支持
E10	女	博士	管理学	科研人员	3~6年	经常	非常支持
E11	女	博士	军事学	教师	10~20年	经常	非常支持
E12	男	博士	工学	科研人员	6~10年	经常	非常支持
E13	女	博士	管理学	教师	3~6年	经常	非常支持
E14	女	硕士	医学	临床工作者	1~3年	经常	非常支持
E15	女	学士	经济学	教师	20年及以上	一般	非常支持
E16	男	博士	医学	医务工作者	10~20年	经常	支持

调查问卷第二部分涉及影响因素指标释义与打分。问卷给出各影响因素及其释义,请被调查者进行打分,打分标准为0~5分。0分表示该影响因素对科学数据引用标准化没有影响,5分表示影响程度最大。选择0分后,在问卷的第三部分可忽略该因素对其他因素的影响关系判断。

调查问卷第三部分涉及影响因素之间的作用机理。被调查者根据各个影响因素的作用大小,将它们两两配对,进行比较,记录关联判断(如果矩阵的行项因素 $S_i$ 会对列项因素 $S_j$ 产生直接的作用,则将 $S_i$ 记录为1;如果矩阵的行项因素 $S_i$ 不会对列项因素 $S_j$ 产生直接的作用,则将 $S_i$ 定义为0)。

## 2.2 科学数据引用标准化影响因素指标体系的确立

基于上述调查结果,展开数据整合与分析。当对影响因素间关联的判断出现较大偏差时,进行二次意见征询,以确认调查结果的客观性,降低误判率。最终以80%为判定依据判断影响因素相关性。调查结果显示,科学数据的价值更倾向于通过数据质量和影响力反映,引文支持可具体在引文创建工具或技术、行为指南等指标中客观体现。

据此,对数据价值、引文支持等部分因素进行了修正,最终确定科学数据引用标准化的18个影响因素,并

将科学数据引用标准化影响因素归纳为科学数据的资源特性、科学数据引用行为、技术环境因素、相关利益主体因素、政策制度因素5个层面,得到最终的科学数据引用标准化影响因素指标体系(见表2)。

## 3 基于ISM-MICMAC的科学数据引用标准化影响因素作用机理分析

### 3.1 构建科学数据引用标准化影响因素的ISM

#### 3.1.1 确定邻接矩阵与可达矩阵

根据ISM,确立18个影响因素的相互关系,结合专家意见,得出邻接矩阵 $A$ (见图1),表达各影响因素间的直接作用关系。

可达矩阵根据布尔矩阵算法规则,将有向图中节点间的路径联系以矩阵的方式表示出来。若 $S_i$ 有路径可达 $S_j$ ,那么 $S_i$ 就能达到 $S_j$ ,例如 $S_1$ 对于 $S_2$ 有直接作用且 $S_2$ 对 $S_{10}$ 有直接作用,则 $S_1$ 与 $S_{10}$ 之间可以建立通路,即有间接影响,记录为1; $S_1$ 和 $S_3$ 不能通过任一路径形成通路,记录为0。按照上述算法原理,利用MATLAB软件对邻接矩阵进行计算,得出可达矩阵 $R$ (见图2),确立各影响因素间的间接作用关系。

表2 科学数据引用标准化影响因素指标体系

层面	影响因素	指标释义	指标来源
科学数据的资源特性	$S_1$ 数据的质量	科学数据完整性、可获得性、时效性的综合体现, 数据来源的可信度	文献[3]、[4]、[5]、[6]、[7]、[8]
	$S_2$ 数据的影响力	科学数据成果的传播扩散和交流能力, 包括内容传播范围和半衰期两个维度	文献[9]
	$S_3$ 数据形式特征	科学数据在生产、传播和使用过程中的形式特征的变化	文献[1]、[10]、[11]
科学数据引用行为	$S_4$ 引用行为偏好	数据使用者在引用数据内容、格式、引文风格等方面固有的习惯和偏好	文献[12]、[13]; 专家意见
	$S_5$ 引用数据类型	被引数据按内容和载体特征主要分为统计整理类、原始记录类和研究成果类, 按描述类型分为文字型、图片、表格型等	文献[13]、[14]
	$S_6$ 引用目的	背景解释、充当论据、补充论证等	文献[13]
	$S_7$ 引文标注格式	科学数据引文标注格式主要包括文后参考文献、图表下方注释引用、正文直接标注引用和文后(非参考文献部分)注释引用等	文献[15]; 专家意见
	$S_8$ 引文创建方式	引文创建方式主要分为手动引文创建、半自动引文创建、自动引文创建	文献[10]
技术环境因素	$S_9$ 引文创建工具或技术	引文创建所需的关键技术, 如DOI Citation Formatter、Zotero等	文献[14]、[16]
	$S_{10}$ 引文管理系统或平台	提供科学数据引文管理、引文要素提取指导的嵌入式或独立平台	文献[17]、[18]、[19]
相关利益主体因素	$S_{11}$ 政府职能部门	主要的科研资助机构, 支持受公共财政资助的科学数据研究成果效益最大化的责任主体	文献[20]
	$S_{12}$ 数据存储/服务机构	致力于科学数据成果的存储、利用和传播的机构, 如图书馆、数据中心等	文献[21]; 专家意见
	$S_{13}$ 数据出版机构	科学数据及其集合的出版、发行、版权管理机构	文献[22]
	$S_{14}$ 科研群体/终端用户	科学数据资源的创建者和使用者, 致力于科学研究和数据交流	文献[1]; 专家意见
政策制度因素	$S_{15}$ 政策制度	由政府职能部门、科学数据产权管理机构制定、发布的科学数据引用政策制度等, 如《科学数据管理办法》《中华人民共和国数据安全法》等	文献[21]
	$S_{16}$ 行业标准	科研资助机构、行业协会为推动科学数据引用实践发展而确立的相关行业标准, 如《信息技术 科学数据引用》等	文献[23]、[24]
	$S_{17}$ 行为指南	科学数据存储及服务机构为规范和引导用户数据引用行为而制定的行为指南, 内容涉及引文创建规则、软件工具操作指南、用户培训等	文献[25]; 专家意见
	$S_{18}$ 学术规范	科研实践部门制定、发布的学术研究、写作规范	文献[26]、[27]、[28]

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

图1 邻接矩阵A

### 3.1.2 科学数据引用标准化影响因素的层次划分

通过对可达矩阵的层层分解, 构建科学数据引用标准化影响因素的分层图谱。根据ISM的构建原理, 定义可达集 $M(S_i)$ 的范围, 即 $S_i$ 可以到达的全部范围, 以

及先行集 $N(S_i)$ , 即与 $S_i$ 有联系的全部影响因素; 可达集 $M$ 和先行集 $N$ 之间的交集 $T$ 代表能够同时受到 $S_i$ 影响并作用于 $S_i$ 的因素。如果可达集 $M(S_i)$ 等于交集 $T$ , 那么 $S_i$ 就属于第一层次 $L_1$ 。剔除 $S_i$ 所属可达矩阵的行和列, 即达第二层次 $L_2$ 。以此类推, 直到最终层次 $L_n$ 。第一层

$$R = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

图2 可达矩阵R

次可达集、先行集与交集见表3。

由表3所示的第一层次可达集、先行集和交集可以得出当*i*=6、7、9、11、14时，满足*M*=*T*=*M*∩*N*，所以因素*S*<sub>6</sub>、*S*<sub>7</sub>、*S*<sub>9</sub>、*S*<sub>11</sub>、*S*<sub>14</sub>为影响科学数据引用标准化的第一层次因素，即*L*<sub>1</sub>={*S*<sub>6</sub>，*S*<sub>7</sub>，*S*<sub>9</sub>，*S*<sub>11</sub>，*S*<sub>14</sub>}，在之后的可达矩阵中剔除因素*S*<sub>6</sub>、*S*<sub>7</sub>、*S*<sub>9</sub>、*S*<sub>11</sub>、*S*<sub>14</sub>对应的行列，即为第二级层次的分解结果（见表4）。

由表4所示的第二层次可达集、先行集与交集，可

表3 第一层次可达集、先行集与交集

因素	<i>i</i> 的取值		
	<i>M</i> ( <i>S<sub>i</sub></i> )	<i>N</i> ( <i>S<sub>i</sub></i> )	<i>T</i> ( <i>S<sub>i</sub></i> )
<i>S</i> <sub>1</sub>	1, 2, 9~18	1	1
<i>S</i> <sub>2</sub>	2, 9~18	1, 2	2
<i>S</i> <sub>3</sub>	3~9	3	3
<i>S</i> <sub>4</sub>	4, 5, 7~9	3~5, 8	4, 5, 8
<i>S</i> <sub>5</sub>	4, 5, 7~9	3~5, 8	4, 5, 8
<i>S</i> <sub>6</sub>	6	3, 6	6
<i>S</i> <sub>7</sub>	7	3~5, 7, 8	7
<i>S</i> <sub>8</sub>	4, 5, 7~9	3~5, 8	4, 5, 8
<i>S</i> <sub>9</sub>	9	1, 5, 8~10, 12, 13, 15~18	9
<i>S</i> <sub>10</sub>	9~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>11</sub>	11	1, 2, 10~13, 15~18	11
<i>S</i> <sub>12</sub>	9~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>13</sub>	9~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>14</sub>	14	1, 2, 10, 12~18	14
<i>S</i> <sub>15</sub>	9~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>16</sub>	9~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>17</sub>	9~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>18</sub>	9~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18

表4 第二层次可达集、先行集与交集

因素	<i>i</i> 的取值		
	<i>M</i> ( <i>S<sub>i</sub></i> )	<i>N</i> ( <i>S<sub>i</sub></i> )	<i>T</i> ( <i>S<sub>i</sub></i> )
<i>S</i> <sub>1</sub>	1, 2, 10, 12, 13, 15~18	1	1
<i>S</i> <sub>2</sub>	2, 10, 12, 13, 15~18	1, 2	2
<i>S</i> <sub>3</sub>	3~5, 8	3	3
<i>S</i> <sub>4</sub>	4, 5, 8	3~5, 8	4, 5, 8
<i>S</i> <sub>5</sub>	4, 5, 8	3, 4, 5, 8	4, 5, 8
<i>S</i> <sub>8</sub>	4, 5, 8	3, 4, 5, 8	4, 5, 8
<i>S</i> <sub>10</sub>	10, 12, 13, 15~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>12</sub>	10, 12, 13, 15~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>13</sub>	10, 12, 13, 15~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>15</sub>	10, 12, 13, 15~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>16</sub>	10, 12, 13, 15~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>17</sub>	10, 12, 13, 15~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18
<i>S</i> <sub>18</sub>	10, 12, 13, 15~18	1, 2, 10, 12, 13, 15~18	10, 12, 13, 15~18

以看出当*i*=4、5、8、10、12、13、15、16、17、18时，满足*M*=*T*=*M*∩*N*，所以因素*S*<sub>4</sub>、*S*<sub>5</sub>、*S*<sub>8</sub>、*S*<sub>10</sub>、*S*<sub>12</sub>、*S*<sub>13</sub>、*S*<sub>15</sub>、*S*<sub>16</sub>、*S*<sub>17</sub>、*S*<sub>18</sub>为影响科学数据引用标准化的第二层次因素，即*L*<sub>2</sub>={*S*<sub>4</sub>，*S*<sub>5</sub>，*S*<sub>8</sub>，*S*<sub>10</sub>，*S*<sub>12</sub>，*S*<sub>13</sub>，*S*<sub>15</sub>，*S*<sub>16</sub>，*S*<sub>17</sub>，*S*<sub>18</sub>}。在可达矩阵中剔除相应的行和列，即为第三层次的分解结果（见表5）。

表5 第三层次可达集、先行集与交集

因素	<i>i</i> 的取值		
	<i>M</i> ( <i>S<sub>i</sub></i> )	<i>N</i> ( <i>S<sub>i</sub></i> )	<i>T</i> ( <i>S<sub>i</sub></i> )
<i>S</i> <sub>1</sub>	1, 2	1	1
<i>S</i> <sub>2</sub>	2	1, 2	2
<i>S</i> <sub>3</sub>	3	3	3

由表5所示的第三层次可达集、先行集与交集，可以看出当 $i=2, 3$ 时，满足 $M=T=M \cap N$ ，所以 $S_2, S_3$ 为影响科学数据引用标准化的第三层次因素，即 $L_3 = \{S_2, S_3\}$ 。在可达矩阵中剔除相应的行和列，即为第四层次的分解结果（见表6）。

表6 第四层次可达集、先行集与交集

因素	i的取值		
	$M(S_i)$	$N(S_i)$	$T(S_i)$
$S_1$	1	1	1

由表6所示的第四层次可达集、先行集与交集，可以看出当 $i=1$ 时，满足 $M=T=M \cap N$ ，所以 $S_1$ 为第四层次因素，即 $L_4 = \{S_1\}$ 。

对可达矩阵 $R$ 中的各行、各列按影响因素层次划分，建立矩阵 $R'$ （见图3），并对其进行三角分块处理，构成单位矩阵。 $R'$ 中的1指的是邻近层次的影响因素之间存在相互作用关系。

基于不同层次影响因素间的作用关系，确定科学数据引用标准化影响因素的层级结构（见图4）。科学数据引用标准化的影响因素分为顶层因素、中层因素

	$S_6$	$S_7$	$S_9$	$S_{11}$	$S_{14}$	$S_4$	$S_5$	$S_8$	$S_{10}$	$S_{12}$	$S_{13}$	$S_{15}$	$S_{16}$	$S_{17}$	$S_{18}$	$S_2$	$S_3$	$S_1$
$S_6$	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$S_7$	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$S_9$	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$S_{11}$	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$S_{14}$	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
$S_4$	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
$S_5$	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
$S_8$	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
$S_{10}$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0
$S_{12}$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0
$S_{13}$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0
$S_{15}$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0
$S_{16}$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0
$S_{17}$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0
$S_{18}$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0
$S_2$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	0	0
$S_3$	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0
$S_1$	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1

图3 可达矩阵的重新排列 $R'$

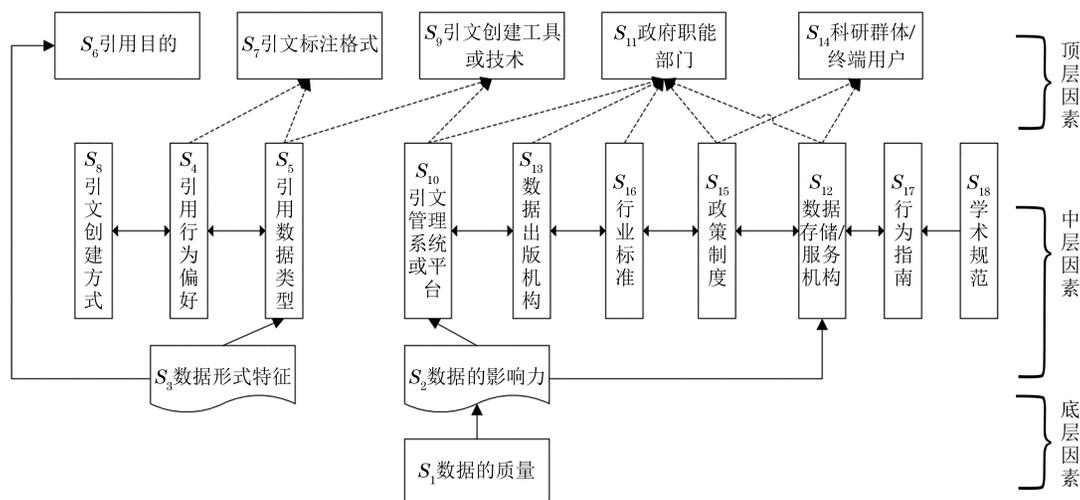


图4 科学数据引用标准化影响因素的层级结构

和底层因素,图4中的箭头指向为影响因素之间的作用关系。

(1)底层因素的分析。数据的质量( $S_1$ )为底层影响因素,对其他各因素产生直接或间接影响,在科学数据引用标准化过程中发挥着主要的驱动效应,是推进科学数据引用标准化进程中需首要关注的因素,这表明数据质量对于鼓励和推进科学数据引用标准化非常关键,应给予足够的重视。

(2)中层因素的分析。①数据形式特征( $S_3$ )、数据的影响力( $S_2$ )是科学数据引用标准化的中层影响因素,由下层影响因素传导至上层影响因素,同时,这些因素直接作用于中层影响因素。其中,数据形式特征( $S_3$ )虽然处于中层,并未受到底层因素的影响,其驱动作用与数据的质量( $S_1$ )相同,需单独求解。②引用行为偏好( $S_4$ )、引用数据类型( $S_5$ )、引文创建方式( $S_6$ )、引文管理系统或平台( $S_{10}$ )、数据存储/服务机构( $S_{12}$ )、数据出版机构( $S_{13}$ )、政策制度( $S_{15}$ )、行业标准( $S_{16}$ )、行为指南( $S_{17}$ )和学术规范( $S_{18}$ )位于模型的中层,直接影响顶层因素。同时引用行为偏好( $S_4$ )、引用数据类型( $S_5$ )、引文创建方式( $S_6$ )三者之间存在直接作用关系,引文管理系统或平台( $S_{10}$ )、数据存储/服务机构( $S_{12}$ )、数据出版机构( $S_{13}$ )、政策制度( $S_{15}$ )、行业标准( $S_{16}$ )、行为指南( $S_{17}$ )六者之间存在直接作用关系。应注意的是,尽管该组因素为中层影响因素,但从分析结果来看,其不受或较少受底层因素影响,需单独求解。

(3)顶层因素的分析。引用目的( $S_6$ )、引文标注格式( $S_7$ )、引文创建工具或技术( $S_9$ )、政府职能部门( $S_{11}$ )和科研群体/终端用户( $S_{14}$ )为顶层因素,对科学数据引用标准化产生直接影响,这些因素最易被辨识,同时受中层和底层因素的影响。应从顶层因素着手推进科学数据引用标准化,规划和制定实施方案。

### 3.2 科学数据引用标准化影响因素的MICMAC分析

运用MICMAC分析方法,对ISM中的各影响因素作具体分析,明确其在科学数据引用标准化过程中的指标特性和实践作用。根据可达矩阵 $R$ ,计算得出 $S_1\sim S_{18}$ 各影响因素的驱动力和依赖性指数,进而得出影响因素的MICMAC分析图(见图5),其中横坐标表示

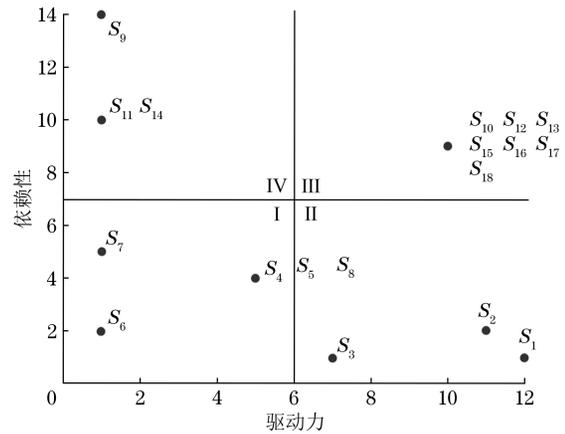


图5 科学数据引用标准化影响因素的MICMAC分析图

驱动力,纵坐标表示依赖性。

第I部落为自治因素,包含引用行为偏好( $S_4$ )、引用数据类型( $S_5$ )、引用目的( $S_6$ )、引文标注格式( $S_7$ )、引文创建方式( $S_8$ )。该组因素的特点是依赖性和驱动力相对较低,即受其他因素影响相对较小,但仍对目标整体(科学数据引用标准化)具有一定影响,因此,在科学数据引用标准化过程中应单独考量。

第II部落为独立因素,包含数据的质量( $S_1$ )、数据的影响力( $S_2$ )、数据形式特征( $S_3$ ),特点是高驱动力、低依赖性。该组因素较少依赖其他因素,但对其他因素有着较大影响,其变化将使不同层次的因素受到影响,对目标整体具有较强的驱动力,这些因素一般列为优先管理因素,在科学数据引用标准化过程中发挥着决定性作用。

第III部落为联系因素,包含引文管理系统或平台( $S_{10}$ )、数据存储/服务机构( $S_{12}$ )、数据出版机构( $S_{13}$ )、政策制度( $S_{15}$ )、行业标准( $S_{16}$ )、行为指南( $S_{17}$ )、学术规范( $S_{18}$ )。该组因素的驱动力和依赖性均较高,因素之间相互作用,共同影响科学数据引用标准化的实现。

第IV部落为依赖因素,包含引文创建工具或技术( $S_9$ )、政府职能部门( $S_{11}$ )和科研群体/终端用户( $S_{14}$ ),特点是高依赖性、低驱动力。该组因素易受体系中其他因素的影响,若要控制这些因素,需从其影响因素入手。因此,在科学数据引用标准化的过程中,应将这些因素列为干预因素,从而有针对性地提出策略。

## 4 基于ISM-MICMAC的科学数据引用标准化推行策略

根据前文分析结果,在明确科学数据引用标准化的关键影响因素、指标特性及作用关系的基础上,有针对性地提出科学数据引用标准化的推行策略。

### 4.1 加强科学数据质量监管和治理,为科学数据引用标准化奠定基础

数据的质量属于底层因素,是影响科学数据引用标准化的关键因素,具有高驱动力、低依赖性的特点。科学数据的共建共享与数据质量密切相关,加强科学数据质量监管是实现科学数据引用标准化的基础。

①政策法规层面,相关部门与政策制定者可通过制定《数据质量法》《数据质量管理条例》《开放政府数据法》等专项法规,并对现有的《科学数据管理办法》和其他相关法律、规定进行修改和完善,为科学数据的质量管理工作提供政策制度保障,并使用强制性和激励性的手段,提高科研人员的科学数据引用意识。②流程监管层面,可将数据质量监管融入科学数据管理全生命周期,明确科学数据质量控制的关键环节,提高对数据质量的控制效率。数据规划阶段,应明确科学数据成果的验收标准、责任主体、灾备管理方案等。数据创建阶段,应关注科研群体的数据素养,规范科研流程。数据组织和存储阶段,确立元数据集,加强期刊与数据库的建设与合作,同时鼓励数据类期刊采用国际引用规范,为后续数据引用提供资源保障<sup>[29]</sup>。对已经取得知识产权的科学数据成果进行标准化操作并实现结构化保存。数据发布和共享阶段,构建科学数据质量评估体系,确定科学数据质量监管的优先级。

### 4.2 落实科学数据引用标准规范,兼顾不同学科引文风格的差异性

落实标准规范对科学数据引用标准化具有直接影响,我国科学数据引用标准化实践的主要制约因素之一一是引用规范与激励政策不完善且执行与宣传工作尚未落实到位,致使数据使用者在数据引用过程中不明确引文创建规则。根据MICMAC分析结果,用户的引用行为偏好、引用数据类型等受其他因素的影响较小,应

单独考量。

与此同时,不同学科的科学数据具有明显的领域特点,应兼顾不同学科引文风格的差异性。不同学科在思维模式、研究方法等方面存在差异,进而在科学数据的引用和共享方式上也不尽相同。如生命科学领域较多引用气象观测数据,且通常会通过文中标注或说明的形式引用。社会科学领域则较多引用政府数据文件,同时采用文后参考列表和文中标注或说明的引用方式。因此,数据管理及相关服务机构可在国家标准《信息技术 科学数据引用》的元数据框架基础上建立符合学科实践的元数据集,明确引文因素和标注规则,为数据使用者创建数据引文提供引导。具体实施过程中,可根据不同学科数据成果类型特点,制定和发布学科科学数据引用标准化实施细则。

### 4.3 依托现有引文技术标准,实现科学数据引文创建与管理的自动化

技术环境因素属于科学数据引用标准化的顶层影响因素,也是实现科学数据引用标准化的“最后一公里”。MICMAC分析结果显示,引文创建工具或技术属于依赖因素,具有高依赖性、低驱动力。与专著、期刊论文等不同,科学数据类型多样,相较于传统文献引用规范而言,科学数据引用规范的制定及实施更加复杂。不同类型的科学数据不仅在内容上存在差异,在载体形式、存储格式等方面也不尽相同,进而在引用格式、引文创建以及引用元素的选取方面存在较大差异,且引用格式由于这些差异的存在而变得难以兼容。因此,在科学数据引用标准化实践中,应重视引文管理平台的建设,提高数据引文创建的便利性。EndNote、Mendeley、Zotero、NoteExpress等在文献引文自动化实践中取得了一定成效。科学数据作为新型文献引用对象,在技术实现路径上与现有参考文献类型具有较高的同质性,现有软件平台的发展可为科学数据引文创建与管理的自动化提供技术支撑。

### 4.4 明确相关利益主体间权责关系,协同推进科学数据引用标准化实践发展

相关利益主体因素是影响科学数据引用标准化的唯一一组能动性因素。科学数据引用标准的实施是一

项复杂的任务工程,涉及多方利益主体,如科研实践团体及科研管理部门、数据存储机构、出版商、图书馆、终端用户等<sup>[30]</sup>。不同主体在科学数据引用标准化中承担着不同的责任,同时秉承着一定的利益诉求,利益主体的非唯一性是推进科学数据引用标准化实践的一个挑战。因此,协同规划、明确各主体间的权责关系对于科学数据引用标准化实践的发展至关重要。具体而言:数据出版机构可基于引用经验和实践,在采纳元数据规范过程中扮演主动角色,在作者、数据生产者、数据存储者、图书馆员和数据发行人之间取得共识,使著者和审稿人能够遵循标准,明确自身的引文需求。数据中心和图书馆可借助平台资源优势,拓展服务内容,在引文创建、引文管理工具使用等方面发挥主要作用,在推进科学数据引用标准化的同时提高馆藏数据资源的利用率和影响力。科研群体既是数据使用者,也是数据成果的主要创建者。从创建者角度,科研人员应积极探索将数据成果纳入科研评价和版权保护体系的路径,规范科研流程,增强数据成果版权意识。从数据使用者角度,科研人员在数据使用过程中应明晰数据来源,积极提升数据引用技能和对现有标准规范的认知水平。

## 5 结语

本文研究科学数据引用标准化影响因素及其作用机理,明确了科学数据引用标准化的影响因素并将其概括为科学数据的资源特性、科学数据引用行为、技术环境因素、相关利益主体因素和政策制度因素5个层面。运用ISM法对科学数据引用标准化的影响因素进行分析,得到影响因素的层级结构图。使用MICMAC法将各因素划分为联系因素、独立因素、自治因素和依赖因素。基于分析结果,提出推进科学数据引用标准化的对策建议,为相关部门有效管理科学数据资源、推进科学数据引用国家标准的实施提供决策参考。不过,本研究也存在一些不足:①基于主观判断提出科学数据引用标准化影响因素的构成,故可能存在缺失影响因素和判断影响因素构成不准确的情况;②问卷调查的对象是专家学者,其主观判断可能会存在偏差,在后续的研究中应扩大调查样本范围来减小误差。

## 参考文献

- [1] 陶蕊,叶继元. 科学数据引用: 源起、问题与研究趋向[J]. 图书馆杂志, 2022, 41(11): 19-29.
- [2] 刘佳. 基于ISM模型的供应链知识流动激励研究[D]. 昆明: 昆明理工大学, 2015.
- [3] HUANG K T, LEE Y W, WANG R Y. Quality information and knowledge[M]. London: Prentice Hall, 1998.
- [4] CAPPIELLO C, FRANCALANCI C, PERNICI B. Data quality assessment from user's perspective[C]//IQIS 04: Proceedings of the 2004 International Workshop on Information Quality in Information Systems. 2004: 68-73.
- [5] AEBI D, PERROCHON L. Towards improving data quality[C]//International Conference on Information Systems & Management of Data. 1999: 273-281.
- [6] MORBEY G. Data quality for decision makers[M]. 2nd ed. Wiesbaden: Springer Gabler, 2013.
- [7] 曹建军,刁兴春,汪挺,等. 数据质量控制研究中若干基本问题[J]. 微计算机信息, 2010, 26(9): 12-14.
- [8] 盛小平,欧阳娟. 国内外开放科学评价研究综述[J]. 情报理论与实践, 2023, 46(11): 174-181.
- [9] Oxford living dictionaries[EB/OL]. [2023-07-10]. <https://en.oxforddictionaries.com/definition/impact>.
- [10] 宋宇,真漆,汤珊红. 数据引用的共同原则[J]. 情报理论与实践, 2015, 38(8): 145.
- [11] 朱艳华,廖方宇,胡良霖,等. 科学数据安全标准规范关键问题探索[J]. 信息安全, 2021, 21(11): 1-8.
- [12] DURYE M, HOCHMAN M, PARFITT A. Measuring the impact of research[EB/OL]. [2023-08-10]. <http://facdent.hku.hk/docs/ResGlob2007.pdf>.
- [13] 邱均平,肖博轩,徐中阳,等. 国内外图书情报领域数据引用特征的多维度分析[J]. 情报理论与实践, 2022, 45(9): 44-50, 36.
- [14] 张颖. NoteFirst软件在专题文献管理中的应用[J]. 图书馆研究与工作, 2014(2): 62-64.
- [15] 司莉,伍丹. 我国图书情报领域科学数据引用行为调查与分析[J]. 情报科学, 2020, 38(2): 11-16.
- [16] 涂志芳. 科学数据出版的基础问题综述与关键问题识别[J]. 图书馆, 2018(6): 86-92, 100.
- [17] KOTARSKI R, RELLY S, SCHRIMPF S, et al. Report on best practices for citability of data and on evolving roles in scholarly communication[EB/OL]. [2023-07-31]. <http://rep4-vm.awi>.

- de/31396/1/ODE-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf.
- [18] 王翠萍, 宋雯琪, 姜鑫妍, 等. 我国科学数据共享平台数据描述及引用规范调查研究[J]. 图书馆学研究, 2022 (11): 76-85.
- [19] 李建霞, 余丹丹. 科学数据共享平台用户感知有用性分析: 以国家科技基础条件平台为例[J/OL]. 情报杂志: 1-7[2023-06-04]. <https://kns.cnki.net/kcms/detail/61.1167.G3.20230602.1038.016.html>.
- [20] 崔海媛, 聂华, 吴越, 等. 公共资助机构开放获取政策研究与实施: 以国家自然科学基金委员会基础研究知识库开放获取政策为例[J]. 大学图书馆学报, 2017, 35 (3): 79-86.
- [21] 史雅莉, 赵雪芹. 合作博弈视角下科学数据引用主体间的关系探析[J]. 数字图书馆论坛, 2019 (1): 15-20.
- [22] 吴立宗, 王亮绪, 南卓铜, 等. 科学数据出版现状及其体系框架[J]. 遥感技术与应用, 2013, 28 (3): 383-390.
- [23] 张静蓓, 吕俊生, 田野. 国外科学数据引用研究进展[J]. 图书情报工作, 2014, 58 (8): 91-95, 138.
- [24] 王丹丹. 科学数据规范引用关键问题探析[J]. 图书情报工作, 2015, 59 (8): 42-47, 53.
- [25] 邢文明, 洪程. 开放为常态, 不开放为例外: 解读《科学数据管理办法》中的科学数据共享与利用[J]. 图书馆论坛, 2019, 39 (1): 117-124.
- [26] 张贵兰, 王健, 周国民, 等. 科学数据用户相关性标准研究[J]. 图书情报工作, 2019, 63 (4): 112-121.
- [27] 司莉, 贾欢. 科学数据的标准规范体系框架研究[J]. 图书馆, 2016 (5): 5-9.
- [28] 朱艳华, 高瑜蔚, 胡良霖, 等. 我国科学数据标准规范实践与思考[J]. 中国科学数据, 2023, 8 (1): 165-174.
- [29] 戚筠, 何琳. 数据生命周期视角下国外人文社会科学领域数据期刊政策及启示[J]. 中国科技期刊研究, 2023, 34 (3): 315-324.
- [30] 张丽丽, 黎建辉. 数据引用的相关利益者分析[J]. 情报理论与实践, 2014, 37 (7): 44-47.

## 作者简介

史雅莉, 女, 博士, 副教授, 研究方向: 信息组织、科学数据管理、数据引用、数据伦理, E-mail: syl19890729@163.com。  
沈心悦, 女, 硕士研究生, 研究方向: 科学数据管理、语义关联。  
孙蕴哲, 女, 硕士研究生, 研究方向: 信息组织、科学数据引用。

### Influencing Factors and Mechanism of Scientific Data Citation Standardization Based on ISM-MICMAC

SHI YaLi SHEN XinYue SUN YunZhe  
(School of History and Culture, Hubei University, Wuhan 430062, P. R. China)

Abstract: The arrival of data intensive paradigms has made data play an increasingly important role as a strategic resource in scientific research and various industries. It is particularly important to systematically explore the internal and external influencing factors of scientific data citation standardization and promote the development of scientific data citation standardization practices. The article uses the ISM-MICMAC analysis method to analyze the indicator characteristics of each influencing factor and the mechanism of action between different influencing factors, and constructs a hierarchical relationship structure of influencing factors for scientific data citation standardization. Based on this, promotion strategies for the standardization of scientific data citation are proposed: strengthening the supervision and governance of scientific data quality, laying the foundation for the standardization of scientific data citation; implementing scientific data citation standards and norms, taking the differences in citation styles among different disciplines into account; relying on existing citation technology standards, achieving automation of scientific data citation creation and management; clarifying the rights and responsibilities among relevant stakeholders, and promoting the development of standardized scientific data citation practices together.

Keywords: Scientific Data; Scientific Data Citation; Standardization; Influencing Factor

(责任编辑: 王玮)