

·信息素养与教育·

嵌入科研 workflow 与数据生命周期的数据素养能力研究

胡卉^{1,2} 吴鸣¹

(1.中国科学院文献情报中心 北京 100190)

(2.中国科学院大学 北京 100049)

摘要:文章首先分析归纳了科研 workflow 与数据生命周期的一般流程,接着调研数据素养核心能力研究现状,剖析国内外学者提出的几种较有代表性的数据素养核心能力体系,最后梳理了嵌入科研 workflow 和数据生命周期的数据素养核心内容体系,绘制了研究人员应具备的数据素养能力框架。

关键词:数据素养;科研 workflow;科研数据生命周期

中图分类号:G254.97 文献标识码:A DOI:10.11968/tsyqb.1003-6938.2016085

Core Competencies of Data Literacy Embedded in Research Workflow and Data Lifecycle

Abstract The research work flow processes and general research data lifecycle are summarized at first, and then research status of data literacy core competencies is investigated and several representative core competencies frameworks are analyzed, and finally core competencies framework of data literacy around research work flow processes and research data lifecycle are built.

Key words data literacy; research workflow; research data lifecycle

1 引言

大数据时代的到来正在改变 21 世纪公民的素养格局,尤其是数据密集型第四科研范式下,具备良好的数据素养和熟练的数据技能已成为研究群体开展研究工作的重要能力保障^[1]。然而,实践中庞大的数据量和复杂的数据管理问题使得研究人员面临严峻挑战,为适应数据驱动的新型研究环境,应对科研实践中的各种数据问题,培养研究人员的数据素养能力势在必行。那么数据素养是什么? M.Schild^[2]认为数据素养是根据科研需求在不同数据源中获取、操作和总结数据,并从中推断结论的一种能力;J. Qin 和 D.Ignazio^[3]认为数据素养是“研究者在科研过程中收集、处理、操作、评估和利用数据的能力”;P. Calzada 和 M.A.Marzal^[4]认为数据素养“包含个人获取、理解、批判性地评估和管理数据的能力,并在使用数据的过程中遵守道德规范”。总体而言,数据素

养强调的是一种正当地发现和获取数据、批判地选择和评估数据、规范地管理和处理数据、合理地利用和共享数据的意识和能力。

目前,国内外学者已就数据素养能力体系展开探索,并积累了一定的成果和经验,主要集中在三个方面:一是从数据素养的概念和内涵出发,梳理数据素养包含的各种能力,如 J.Qin^[3]、郝媛玲和沈婷婷^[5]等;二是图书馆员在长期从事科研数据管理服务的基础上,概括和总结数据管理过程中所需的技能,如 Carlson.J^[6]、A.L.Ogier^[7]等;三是从数据素养的上位类概念信息素养出发,在信息素养能力标准中提取数据相关的能力,如 C. P.Javier 和 M.A. Marzal^[4]等。然而,从研究项目启动之初的数据产生、项目实施中的数据处理和分析到项目结题时基于数据得出研究结论、发表科研成果,再到后续研究中对数据的二次开发和利用,科研数据已成为科学研究工作的基本单元渗透到科研 workflow 的各个环节,数据素养与科研

收稿日期 2016-05-27,责任编辑 魏志鹏

工作流、科研数据生命周期密不可分。因此,构建嵌入科研 workflow、围绕科研数据生命周期的数据素养能力框架,对于更好地培养研究人员的数据意识和数据管理技能具有重要意义。基于此,本文从科研 workflow 和数据生命周期的一般流程出发,调研了数据素养核心能力研究现状,整理并分析了国内外学者提出的几种典型的数据素养核心能力体系,梳理并归纳了嵌入科研 workflow 和数据生命周期的数据素养核心内容体系,绘制研究人员应具备的数据素养技能流程图,以期为衡量研究人员的数据素养能力和探索数据素养教育实践提供参考和建议。

2 科研 workflow 与数据生命周期

数据从收集(或形成)、加工、保存、传播、检索、存取与利用,到消失或不再被利用的过程,就是数据的生命周期^[8]。随着科技的发展和时代的快速进步,有些数据永久地休眠,再也得不到利用;而很大一部分数据却只是进入了暂时的休眠状态,它们将来可能会根据人类的需要再次被激活和更新,进入下一个“生命周期”,周而复始,生生不息。在数字科研环境下,“生命周期”不同于产生到消亡的“生命期”,它是一个循环过程,经过数字资源管理和长期保存,实现数字资源发现和再利用^[9]。在科研实践过程中,处于生命周期不同阶段的数据具有显著差异,不论是数据的类型格式还是内容含义,亦或数据的量级特性,均不可一概而论。正确认识科研数据生命周期,把握生命周期各个阶段的数据特性,培养良好的数据素养能力,是实现科研数据规范管理和发挥数据价值最大化的前提和基础。

目前,科研 workflow 与数据生命周期已得到业界的广泛重视,英国联合信息系统委员会(JISC)、美国地质调查局(USGS)等国家组织委员会,英国数据管理中心(DCC)、英国数据档案(UKDA)、地球数据观察网络(DataONE)等数据管理专业机构以及弗吉尼亚大学、阿尔伯塔大学、昆士兰大学等学术研究机构纷纷就科研数据生命周期展开探索研究,建立科研 workflow 与数据生命周期模型,以规范和指导数据管理工作的正常开展。

为深入了解科研实践的具体流程,笔者选取了

国际上较典型的科研数据生命周期进行剖析,以总结和梳理科研 workflow 和数据生命周期的一般流程。8种模型具体为:英国高等教育、继续教育和技能培训提供研究支撑的英国联合信息系统委员会(JISC)^[10];英国专门负责大型社会和经济数据的采集、监管并提供数据访问与获取的专业机构英国数据档案(UKDA)^[11];世界领先的数字信息管理中心英国数字内容管理中心(DCC)^[12];世界上最大的社会科学数据中心美国校际社会科学数据共享联盟(ICPSR)^[13];地球环境科学领域的管理专业组织地球数据观察网络(DataONE)^[14];专注于统计数据和社会科学数据描述的国际标准协会数据文档倡议联盟(DDI)^[15];美国弗吉尼亚大学图书馆^[16];加拿大阿尔伯塔大学图书馆^[17](见表1)。

笔者在分析梳理国外典型科研 workflow 与数据生命周期模型的基础上,结合国内研究实践,绘制了嵌入科研 workflow 的科研数据生命周期一般流程(见图1)。除了项目启动、项目实施和项目结项三个核心阶段外,还包括项目启动前的研究构想以及项目结束后的后续研究。科研数据生命周期的一般流程包括八个阶段,分别是数据管理计划、数据产生与收集、数据管理与组织、数据处理与分析、数据存储、数据发表与共享、数据发现与获取、数据再利用,各个阶段环环相扣,紧密相连。下文将以此为基础梳理和构建围绕科研 workflow 和数据生命周期的数据素养能力框架。

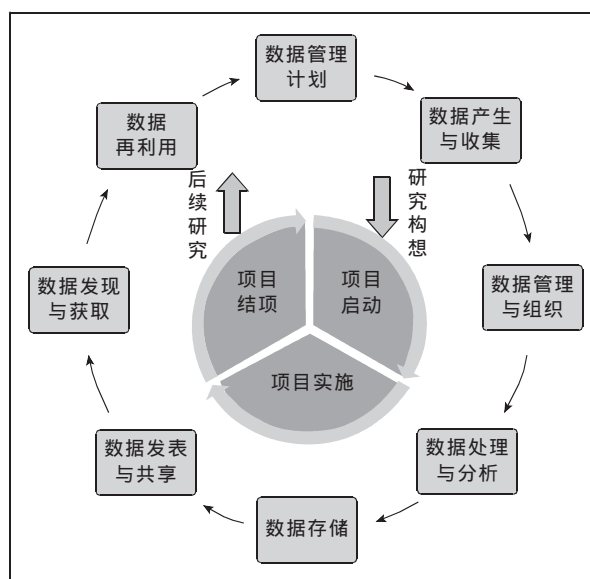


图1 嵌入科研 workflow 的科研数据生命周期一般流程

表1 8种科研工作流与数据生命周期模型

| 机构名称 | 科研工作流与数据生命周期模型 |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| JISC | 分为科研生命周期和数据生命周期,科研生命周期包括研究灵感、组建研究团队、撰写项目书、研究过程和研究成果发表五个阶段;数据生命周期嵌入研究过程,包括数据产生、数据管理、数据分析和数据共享四个环节 |
| UKDA | 科研数据生命周期划分为六个阶段:第一阶段创建数据(研究设计;数据管理计划;数据共享协议;查找现有数据;收集数据;收割或创建元数据);第二阶段处理数据(数据输入、数字化、转化;数据检查、验证、清晰、必要的匿名化处理、数据描述、管理与存储数据);第三阶段分析数据(数据解释;数据导出;得出研究结论;作者出版;数据存储准备);第四阶段保存数据(将数据转化为最佳格式;将数据迁移到最佳存储介质;数据备份与保存;创建元数据文档和记录文档;数据存档);第五阶段共享数据(数据传播;数据共享;控制访问;数据知识产权;数据推广);第六阶段数据再利用(后续研究;新的研究;研究验证与回顾;审查结果;用于教学) |
| DCC | 数据监管生命周期模型包括两个维度:外维是科研数据生命周期的整体概貌,分为数据创建与收集、数据评估与选择、数据处理、数据保存与存档、数据访问与再利用和数据转换六个环节;内维模型是数据生命周期的内涵,包含五个层次,层层递进。第一层也是最里层是数据,即数字对象和数据库,数字对象既包括文本文件、图片文件、声音文件及其相关的标识符和元数据等简单数字对象,也包括用以描述其他数字对象产生的复杂对象,如网页,数据库则是存储于计算机的结构化数据;第二层是数据的描述和表达,重视选取恰当的元数据标准,以便清楚完整地揭示数据集及其背景内容;第三层是制定数据存储计划,既包括临时保存措施也包含数据的长期保存计划;第四层是学术交流与传播,即通过发表数据论文、存缴数据仓储以实现数据价值的传播;第五层是随着数据成果的发表,完结的数据应根据数据管理计划进行监管和存储 |
| ICPSR | 科研数据生命周期分为六个阶段,第一阶段是撰写项目申请书和数据管理计划(获取数据归档建议;创建数据管理计划以确保数据资源的长期可用性);第二阶段是项目启动(确定数据的格式和内容;对研究数据和方法进行预测或抽样测试);第三阶段是数据收集和创建(遵循最佳实践;在数据方面确定数据的完整性、数据的变量名称、数据标签、数据分组、数据编码和缺失数据;在文件组织方面使用DDI的元数据标准);第四阶段是数据分析(数据集和数据文件的管理;建立合理的文件结构;备份数据和文档);第五阶段是数据共享准备(明确数据公开的风险,设置数据存储的格式;确定数据存储地点);第六阶段是数据存档(将数据打包完整存档;遵守数据共享和传播协议) |
| DataONE | 科研数据周期包括八个阶段,分别是数据管理计划、数据收集、质量控制、数据描述、数据存储、数据发现、数据集成、数据分析 |
| DDI | 科研数据周期包括八个阶段研究主题、数据收集、数据处理、数据存档、数据传播、数据发现、数据分析和数据再利用 |
| 弗吉尼亚大学图书馆 | 科研工作流与数据生命周期集成模型包括项目申请书撰写、项目启动、数据收集、数据分析、数据共享、项目结题、数据存档、数据发现、数据再利用九个阶段,其中数据生命周期指从数据收集到数据再利用的过程 |
| 阿尔伯塔大学图书馆 | 科研工作流与数据生命周期融合,分为研究设计、数据收集、数据处理、数据访问与传播、数据分析、形成研究成果、数据发现和数据再利用八个阶段 |

3 数据素养核心能力研究现状

能力标准为图书馆服务开展和各级各类教学提供了明确的目标和方向,也为各个机构的课程设计提供了基本的框架和指南^[18]。数据素养是信息素养的子集,目前,信息素养领域已建立了较完善的能力标准体系,既有以美国、英国、澳大利亚、加拿大等区域性信息素养能力标准,又有面向中小学教育、高等教育的层级性信息素养能力标准,以及大众传播学、人类与社会科学、护理学、科学与工程技术等以学科划分的领域信息素养能力标准。尽管这些标准对于数据素养评估和教学有一定的参考意义,但就数据素养应用实践而言,仍亟需数据素养能力标准或框架来衡量研究人员的数据素养能力和指导数据素养教育工作的开展。

目前,这一问题已获得国内外学者的广泛关注,他们从不同的角度,结合数据管理服务和数据素养

教育探索的实践经验,构建数据素养能力体系,推动数据管理培训和数据素养教育的发展。如美国雪城大学 J.Qin 等^[3]对比了信息素养、数字素养和科学数据素养的不同概念,指出科学数据素养包含数据收集、数据处理、数据管理、数据评估和数据利用五个方面的技巧和能力。并以此为基础,为学校 STEM 各专业本科生开设科学数据素养课程;美国弗吉尼亚理工大学图书馆为满足学校师生的数据管理需求, A.L.Ogier 等^[7]图书馆员开发了数据素养课程,其教育内容模型包括八个方面:数据管理与组织、数据转换与互操作性、数据共享与获取、元数据与质量控制、数据科学、数据监管、数据存档与保存、数据伦理与责任;美国印第安纳大学图书馆 S.Konkiel 等^[9]设计了包含 10 种能力的数据素养核心能力框架:数据管理与组织、数据格式与数据类型、数据可视化、数据质量控制、数据发现与获取、数据分析、元数据、数据伦理、数据保存、数据复用,将数据管理素养集成

到数据可视化课程中,根据学生的数据技能需求,有针对性地制定学习目标和评估方案;巴西里约热内卢联邦大学 A.Tygel 和 R.Kirsch^[20]基于教育学家 P. Freire 的素养模型定义了数据素养核心能力,包括数据读取(数据意识、数据发现与获取、数据收集)、数据处理(数据处理技术、数据处理工具)、数据交流(数据类型、数据格式、数据评估、数据伦理)和数据生产(数据格式、数据发表工具)四种核心能力,用以指导和规范图书馆的数据管理课程;西班牙卡洛斯三世大学 P.Javier 和 M.A.Marzal 等^[4]从信息素养能力标准出发,选取美国学校图书馆员协会(AASL)制定的《共同核心州立标准》、美国大学与研究图书馆协会(ACRL)制定的《高等教育信息素养能力标准》、《科学与工程信息技术信息素养能力标准》、《人类学与社会科学信息素养能力标准》、《护理学信息素养能力标准》、《政治学信息素养能力标准》,梳理其中有关数据管理的内容,结合图书馆科学数据管理服务,构建了包含数据意识(什么是数据、数据背景)、数据发现和获取(数据来源、数据获取方法)、数据阅读理解和评价(数据阅读与表达、数据评估)、数据管理(数据和元数据的收集与管理)、数据利用(数据操作、数据分析、数据使用伦理)五大指标的数据素养核心能力框架;瑞士 R.Schneider^[21]在“全球信息素养研究与实践的共性与挑战”大会上指出科研数据素养和数据管理能力包括 8 个方面的内容,分别是数据识别、数据处理、数据计划、数据存储、数据保护、数据评估、数据管理、数据交流。国内学者也在构建数据素养能力体系方面取得了一定的研究成果。如上海大学的郝媛玲和沈婷婷^[5]认为数据素养的核心内容包含五个方面的基本要素,分别是数据态度(数据道德观、数据价值观)、数据意识(主体意识、获取意识、共享意识、更新意识、安全意识、人才意识)、数据知识(特点与类型、作用于效应、规律与规范、转换方式、搜索方法)、数据技能(获取、处理、利用、展示、评价、再创造)和数据伦理(法制观念、数据安全、数据道德、知识产权),旨在为我国建立数据素养培养机制提供建议。

为了更深入展示国内外数据素养能力体系研究现状,笔者选取了美国普渡大学图书馆研究团队、加

拿大达尔豪斯大学图书馆研究团队和国内华东师范大学团队构建的数据素养核心能力框架作为案例进行详细介绍。

(1)普渡大学图书馆

普渡大学、康奈尔大学、明尼苏达大学和俄勒冈大学四校图书馆联合开展数据信息素养教育项目(DIL),以 Carlson.J 为代表的研究团队通过剖析 ACRL《高等教育信息素养标准》,访谈调研教师和学生数据信息素养需求,并总结地理信息学等图书馆课程经验,构建了包含 12 项核心能力的素养框架体系:数据实践规范、数据转换与互操作、数据监管和再利用、数据管理和组织、数据保存、数据处理和分析、数据质量和记录、数据可视化和表示、数据库和数据格式、数据发现和获取、元数据和数据描述、数据伦理与道德(见表 2)。DIL 项目组在该能力框架的基础上,调研自然资源、土木工程、计算机科学等不同领域研究群体的数据需求,根据需求设计课程大纲展开针对性培训和教育,提高研究人员的数据管理能力^[6]。

(2)加拿大达尔豪斯大学图书馆

加拿大达尔豪斯大学图书馆的 C. Ridsdale 等^[22]学者在数据管理服务经验的基础上梳理出科研人员在数据管理过程中所需的数据素养能力,构建了包括数据管理概念、数据收集、数据管理、数据评估和数据应用五方面的数据素养能力矩阵(见表 3),拟在能力矩阵的基础上设计数据素养课程,帮助学校师生员工提高数据管理技能。

(3)华东师范大学隆茜

我国华东师范大学的隆茜^[23]在综合国内外文献对数据素养内涵、组成能力的阐述的基础上,从数据意识、数据获取能力、数据处理与分析能力、数据交流能力、数据评价能力和数据道德六个维度构建了高校师生数据素养能力评价指标体系(见表 4)。并以此为基础编制数据素养能力调查问卷,分析高校师生的数据素养能力现状,为数据素养教育奠定基础。

笔者对以上 10 种数据素养能力框架进行梳理与解析,深入剖析能力框架的构成要素,共罗列了 21 种数据素养技能,分别是数据意识、数据收集与创建、数据转换与互操作、数据管理与组织、数据保存、

表2 DIL 项目开发的数据信息素养能力框架及其相关的数据技能

| 数据信息素养核心能力 | 与数据信息素养能力相关的数据技能 |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 数据实践规范 | 了解特定学科领域及其子领域内的数据管理、数据共享、数据监管和数据存储的规范和实践情况;了解相关领域的数据标准,如元数据、数据质量、数据格式,以及如何运用这些数据标准 |
| 数据转换与互操作 | 能将数据在不同格式之间灵活转换;了解数据转换时的潜在风险和损失;了解将数据转换成标准格式后在数据再利用等方面的优点 |
| 数据监管和再利用 | 了解数据除了原始利用目的外还有其他作用,如支持研究验证和他人的再利用;能够区分哪些数据可用于未来他人的再利用;了解尽管管理数据是一项复杂的、耗费时间精力的工作,但对于数字科研环境下的合作研究至关重要;数据管理应根据数据生命周期指定数据管理计划;明确数据管理的各个环节和需求,并能具体落实到项目实践中;了解如何引用数据以及创建数据时如何让自己的数据可引用 |
| 数据管理和组织 | 了解数据生命周期,学会创建数据管理计划,并根据数据管理落实到具体的数据管理工作,对原始数据集进行处理;学会创建标准的数据管理流程文档 |
| 数据保存 | 了解数据保存的意义及相关成本;了解数据保存的技术、资源以及组成部分;在保存数据时参考其他的数据存储最佳实践;了解数据长期保存对于自己和他人的价值,并选择合理的保存时限;了解数据保存政策并能理解政策的主要内容 |
| 数据处理和分析 | 熟悉领域内数据处理的基本流程以及数据分析工具和技术;了解数据分析工具的作用;使用合适的工作流管理工具来完成重复的数据分析工作 |
| 数据质量和记录 | 识别、记录并解决数据作假、数据缺失和数据错误问题;使用元数据识别数据集潜在的问题;完整记录数据以确保数据的可重复性和他人对数据的可利用性;注明数据的出处,详细记录数据的版本信息 |
| 数据可视化和表示 | 熟练运用学科领域内基本的数据可视化工具;避免对数据图表的误解和错误解释;选择恰当的可视化类型来表示数据,如地图、图表、动画、视频等 |
| 数据库和数据格式 | 了解数据库的基本知识以及如何访问数据库;熟悉领域内的标准数据格式和类型;学会面对不同的研究问题选择恰当的数据格式和类型 |
| 数据发现和获取 | 学会查找和利用学科领域内的数据仓储;学会评估从其他地方获取的数据的质量;除了对数据源进行评估外,还应学会获取、导入和转换数据 |
| 元数据和数据描述 | 理解并学会利用元数据对数据进行描述,使得数据能够被自己和他人正确解读;学会解读元数据,能够正确理解来自外部数据源的数据;理解本体的概念和结构,更好地实现数据共享 |
| 数据伦理和道德 | 了解与数据相关的知识产权、隐私问题和保密问题,以及本领域内数据共享和保存的相关注意事项;学会引用数据或向数据原始作者致谢;在使用数据时避免对数据的误用和错误表达 |

数据处理与分析(技术和工具)、数据质量控制、数据可视化、数据类型、数据格式、数据发现与获取、元数据、数据伦理与道德、数据评估、数据利用与复用、数据共享、数据长期保存、数据安全、数据解读、数据引用、数据交流,各能力框架及其对应的数据素养技能(见表5),这将为本文构建科研数据生命周期的数据素养核心能力框架提供参考和借鉴。

4 嵌入科研工作流与数据生命周期的数据素养能力框架

在考察数据素养核心能力研究现状、总结科研工作流和科研数据生命周期的一般流程的基础上,本文梳理了嵌入科研工作流与数据生命周期的数据素养核心能力框架(见图2)。数据素养能力框架从科研工作流出发,包括项目启动前的数据管理计划;项目实施中的数据产生与收集、数据管理与组织、数据处理与分析、数据存储;项目结题时的数据发表与

共享,以及项目结束后的数据发现与获取、数据再利用等8个生命周期阶段的22项数据管理能力。

(1)数据意识。数据意识是一个抽象的概念,具体来说,作为一名数据工作者,应理解什么是科研数据和为什么要管理科研数据等问题。科研数据是为了特定的目标所收集、观测、创建、分析的信息^[24]。良好的数据管理实践包含但不限于以下优点:提高研究成果的影响力,增加数据引用率;避免重复劳动,节省时间;确保研究成果可验证性和完整性;保证数据安全,降低数据丢失的风险;满足资助机构和学术出版商的数据管理要求;通过数据再利用,促进新的知识发现^[25]。总体而言,数据管理贯穿科研项目始终,良好的科研数据管理实践是知识再生的基础。规范化地管理科研数据,是确保数据长期保存和持久访问的关键,也是未来数据被其他研究人员理解和复用的前提。因此,研究人员在正式着手数据管理工作之前,应具备基本的数据意识。

表 3 数据素养能力矩阵

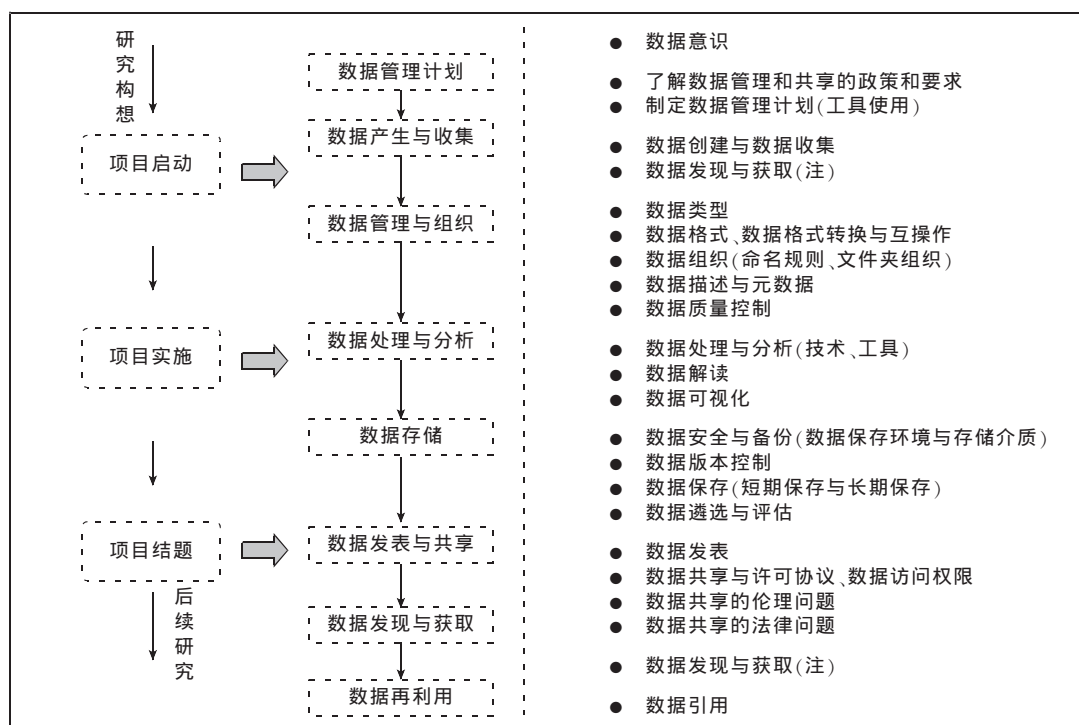
| 知识阈 | 能力指标 | 行动指标 |
|------|----------------|----------------------------------------------------------------------------------|
| 概念框架 | 数据概述 | 理解什么是数据以及数据的相关知识 |
| 数据收集 | 数据发现和收集 | 学会检索数据;识别有用的数据;数据收集 |
| | 数据和数据源质量的评估和选择 | 根据需求批判性地评估数据源;理性地评估数据集质量,识别其中的错误和问题 |
| 数据管理 | 数据组织 | 掌握数据组织方法和工具的基本知识;数据组织需求评估;数据组织 |
| | 数据操作 | 数据清洗方法评估;识别数据的异常值和离散值;数据清洗 |
| | 数据格式转换 | 了解不同数据类型及其转换方法;学会转换数据格式 |
| | 创建和使用元数据 | 创建元数据方案;选取适当的元数据描述数据集 |
| | 数据监管、安全和复用 | 数据管理需求评估,包括保存方案、保存、访问限制、共享要求等;数据安全要求,包括访问权限、存储介质等;数据监管 |
| | 数据存档 | 数据存储需求评估;数据存储的方法和工具评估;数据保存 |
| 数据评估 | 数据工具 | 了解数据分析工具和技术的知识;选择恰当的数据分析工具和技术;数据分析工具和技术应用 |
| | 数据分析基本技能 | 拟定数据分析计划;应用数据分析方法和工具;进行实验数据分析;评估数据分析结果;将数据分析结论与其他发现对比分析 |
| | 数据解释(理解数据) | 数据图片、表格、图示的解析;识别数据中的关键点,分析其内涵;识别数据中的矛盾与偏差 |
| | 利用数据识别问题 | 利用数据找出实践工作中的问题;利用数据识别更高级别的问题,如政策、环境、科学实验、市场、经济等 |
| | 数据可视化 | 以恰当形式进行组织和表达数据;根据数据绘制有意义的图表;评估数据图表的有效性;批判性地评估数据图表的准确度和误差值 |
| | 数据表达(书面) | 对数据分析结果进行解释;受众的需求和主题相关度评估;选择恰当的分析结果呈现形式;选择恰当的表格和可视化结果用以学术交流;清晰明确地展示论点和结论 |
| | 数据支持决策 | 排列从数据抽取信息的优先级;将数据转化成可操作的信息;衡量解决措施和操作方案的优缺点;实施决策和方案 |
| 数据应用 | 批判性思维 | 认识数据中的抽象问题以及数据揭示的其他问题;开展数据工作时批判性地思考 |
| | 数据文化 | 承认数据的重要性;支持并努力创建一个数据学习、研究和决策支持的数据环境 |
| | 数据伦理 | 了解与数据相关的法律和伦理问题;在利用数据开展工作时遵守数据伦理 |
| | 数据引用 | 了解数据引用的背景知识;恰当引用其他数据源的数据 |
| | 数据共享 | 数据共享的平台和方法评估;数据共享的法律和伦理问题 |
| | 数据决策评估 | 后续数据收集以评估先前数据支持决策的有效性;后续数据的分析;将分析结果与其他结论相对比;对根据数据作出的决策和方案进行评估;保留原始结论和决策,实施新决策和方案 |

表 4 高校师生数据素养能力评价指标体系

| 能力维度 | 能力指标 |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------|
| 数据意识 | 能认识数据是科研过程中非常重要的因素 能够认识到科学数据具有生命周期 能够以严谨认真的态度对待与使用科研过程中产生的数据 能够对论文中使用的数据负责 能够保证数据的公正性与开放性 |
| 数据获取能力 | 能够熟练获取各类数据,了解数据源,如获取途径或数据库等?具有检索与收集各种数据的能力 能够对获取的各种形式的数据(格式、类型、特点)做准确 解读 |
| 数据处理与分析能力 | 能够利用软件处理与分析原始数据 能够利用统计分析软件如 SPSS 对获取的数据做恰当的统计 分析 能够结合文献与实际对统计分析结果做准确的解释 |
| 数据交流能力 | 能够正确运用统计量描述数据,如平均值、标准偏差等 能够基于数据统计量准确表述结果 能够恰当采用统计图如柱状图、散点图等来表征并揭示数据中隐含的趋势、变化等 能够利用数据统计量及数据的统计分析结果来支撑论点 能够依据数据做出正确决策 能够利用数据及数据分析结果撰写论文、报告等 |
| 数据评价能力 | 能够审核数据的准确性,剔除错误或无效数据 能够批判性的评价数据,结合实际质疑数据 能够认识数据反映现实具有的局限性,善于结合具体情况 分析数据 |
| 数据道德 | 了解数据采集、使用、分享中所涉及的道德和伦理问题,能够尊重他人的数据,使用时能够注明出处 |

表5 10种数据素养能力框架包含的数据素养技能

| | (1) JacobCarlson | (2) Jian Qin | (3) Ogier A L | (4) Konkiel S | (5) Alan Tygel | (6) Javier Calzada | (7) Ridsdale | (8) SchneiderR | (9) 隆茜 | (10) 郝媛玲 |
|--------------------|---------------------|-----------------|------------------|------------------|-------------------|-----------------------|-----------------|-------------------|-----------|-------------|
| 数据意识 | √ | | | | √ | √ | √ | √ | √ | √ |
| 数据收集与创建 | | √ | | | √ | | √ | | | √ |
| 数据转换与互操作 | √ | | √ | | | | √ | | | √ |
| 数据管理与组织 | √ | √ | √ | √ | | √ | √ | √ | | |
| 数据保存 | √ | | √ | √ | | | | √ | | |
| 数据处理与分析 (技术、工具) | √ | √ | | √ | √ | √ | √ | √ | √ | √ |
| 数据质量控制 | √ | | √ | √ | | | √ | | | |
| 数据可视化 | √ | | | √ | | | √ | | | |
| 数据类型 | | | | √ | √ | | | | | √ |
| 数据格式 | √ | | | √ | √ | | | | | √ |
| 数据发现与获取 | √ | | √ | √ | √ | √ | | | √ | √ |
| 元数据 | √ | | √ | √ | | | √ | | | |
| 数据伦理与道德 | √ | | | √ | √ | √ | √ | | √ | √ |
| 数据评估 | | √ | | | √ | √ | √ | √ | √ | √ |
| 数据利用与复用 | √ | √ | √ | √ | | | | | | √ |
| 数据共享 | | | √ | | | | √ | | | √ |
| 数据长期保存 | | √ | √ | | | | √ | | | |
| 数据安全 | | | | | | | √ | √ | | √ |
| 数据解读 | √ | | | | | √ | | | | √ |
| 数据引用 | √ | √ | | | √ | | √ | | | |
| 数据交流 | | | | | √ | | √ | √ | √ | |



注:在数据产生与收集和发现与获取阶段均涉及数据发现与获取能力。

图2 嵌入科研工作流与数据生命周期的数据素养能力框架

(2) 了解数据管理和共享的政策和要求。随着开放获取事业的推进, 越来越多的资助机构(如美国国家科学基金委 NSF^[26]、美国国立卫生研究院 NIH^[27]、美国航空航天局(NASA)^[28]、英国研究理事会 RCUK^[29]、经济合作与发展组织 OECD^[30]、期刊出版商(如 Nature、Springer、Wiley、Elsevier、RSC、ACS)以及研究机构(如 JISC、ANDS、牛津大学、麻省理工学院、剑桥大学、康奈尔大学)纷纷发布数据管理政策, 要求科研人员提交数据管理计划, 规范科学数据管理流程^[31]。在国内, 2002 年科技部牵头建设“科学数据共享工程”, 目标是整合离散的科学数据资源, 构建面向全社会的网络化、智能化的管理与共享服务体系^[32]。为规范科学数据汇交和共享, 2006 年科技部还制定了“国家科技计划项目科学数据汇交暂行办法”, 规定了汇交科学数据的种类及范围、数据汇交义务人的权利与义务等^[33]。期刊出版商发布了相应的数据政策, 如《现代图书情报技术》杂志要求从 2016 年起, 所有投稿论文需要提交支持论文结论的科学数据, 并通过适当方式供研究共同体或社会公众共享^[34]。在从事科研工作时, 研究人员有必要了解并遵守相应的数据管理政策和规范。

(3) 制定数据管理计划(工具使用)。许多资助机构明确提出了数据管理计划要求, 如 2003 年美国国立卫生研究院发布数据共享政策, 指出“所有向 NIH 申请经费在 500,000 美元以上的科研项目, 建议但不强制提交一份数据管理说明文档, 阐述数据共享的计划和策略, 由于特殊原因不能共享的, 应在说明文档中详细说明^[27]”; 2007 年, 英国生物技术和生物科学研究理事会要求“其资助的科研项目的申请书必须包含数据管理计划, 作为一份不超过 1 页的独立附件, 简要说明数据的管理和共享计划^[35]”; 2011 年, 美国国家科学基金会提出“所有提交到 NSF 的项目申请书必须包含不超过 2 页的名为《数据管理计划》的补充文件, 详细描述该项目将如何根据 NSF 传播和共享科研成果的相关政策开展数据管理工作^[25]”。为此, 国际数据管理组织根据数据管理要求开发了一系列数据管理计划工具, 常用的包括三种: 一是面向美国资助机构要求的由加州大学数字图书馆协会等协作开发的 DMP Tool^[36]; 二是面向英国资助机构要

求的由 DCC 开发的 DMP Online^[37]; 三是面向加拿大资助机构要求的由阿尔伯塔大学开发的 DMP Assistant(曾用名 DMP Builder)^[38]。在申请基金项目时, 选择恰当的数据管理计划工具制定符合资助机构要求的数据管理计划, 已成为申请基金、获得项目资助的重要前提。

(4) 数据创建与数据收集。在项目之初的数据收集阶段, 研究人员需要考虑是否需要创建新的数据集, 或是否可以复用已有的数据源。如果需要创建新的数据, 那么应考虑使用哪些工具、采用哪种方式、应用何种技术来创建和收集数据。

(5) 数据发现与获取。当确定项目需要哪些数据时, 可以通过数据检索判断是否可以复用已有的数据, 这就需要研究人员具备数据发现和获取的技能。数据获取的途径很多, 常规的包括数据中心、数据仓储、机构知识库等, 如英国数据档案中心、澳大利亚国家数据中心等国家数据中心, 生物学 Gen Bank、医药科学 Dryad、地理环境科学 Pangaea 等领域数据仓储, 哈佛大学数据仓储 Harvard Dataverse Network、中国科学院数据云等机构知识库等都支持数据的开放获取。但需要强调的是, 在检索和获取他人数据集时, 应遵守数据复用和数据在分配的限制条件。研究人员在使用来自第三方的数据时, 需要获得许可访问, 在允许范围内使用科研数据, 使用数据引用标准格式注明数据原始来源, 以保证数据的完整性和真实性。

(6) 数据类型。数据类型多种多样, 明确数据类型有利于高效管理数据、提高科研效率。根据数据来源划分, 可将数据分为观测型数据、实验型数据、仿真数据、派生或编译数据、引用或规范数据五种类型^[39](见表 6)。

(7) 数据格式、数据格式转换与互操作。随着技术的发展, 研究人员在存储数据以确保长期访问时, 应充分考虑硬件和软件的存储设施, 选择恰当的数据格式。如文本文件应选择 ODF 格式而不是 Word 格式, 表格文件应选择 ASCII 格式而不是 Excel 格式, 视频文件应选择 MPEG-4 格式而不是 Quicktime 格式, 图片文件应选择 TIFF 或 JPEG2000 格式而不是 GIF 或 JPG 格式, 网页应选择 XML 或 RDF 格式

表6 数据类型及其相关信息

| 数据类型 | 数据产生途径 | 数据类型描述 | 数据实例 |
|---------|----------------------|-----------------------|--------------------------|
| 观测型数据 | 实时捕获的数据 | 通常是不可重复和替代的数据 | 传感器数据、调查数据、样本数据、神经图像等 |
| 实验型数据 | 实验仪器设备产生的数据 | 通常是可以再生的数据,但往往再生成本昂贵 | 基因序列、色谱图、环形线的磁场数据等 |
| 仿真数据 | 由仿真模型生成的数据 | 模型和元数据的输入比输出数据更重要 | 气候模型、经济学模型数据 |
| 派生和编译数据 | 在原始数据集中派生的数据 | 可再生的数据,但是再生成本非常昂贵 | 文本和数据挖掘、编译数据库、3D模型等 |
| 引用或规范数据 | 小型的数据集,通常是发表的预印本中的数据 | 数据可信度高,但通常指一部分或非完整的数据 | 数据银行中的基因序列、化学结构式或空间数据门户等 |

而不是 RDBMS 格式。总体而言,规范并支持格式转换和互操作的数据格式应具备以下特点:非私有的;开放的文档标准;被科研群体普遍使用的数据格式;计算机可读的标准化格式,如 ASCII、Unicode;非加密的;非压缩的^[40]。

(8)数据组织(命名规则、文件夹组织)。在数据完成收集后,应对数据进行规范组织,包括数据和数据集的结构化、制定文件命名规则等。数据结构化时应注意使用文件夹管理数据文件,使用具有揭示意义的词汇来命名文件夹,使用科学的文件夹层次结构,将正在进行的数据文件和已经分析完成的数据文件分开保存,设置最高级别的访问权限等。文件命名指采用一致性的逻辑方式对数据和数据文件夹进行命名,制定文件命名规则时,可参考以下规则:文件名短但具有揭示意义;日期时间使用标准格式;避免使用空格键;避免使用特殊字符;基于主题内容识别数据,而不是仅依靠时间和版本;如果在不同的文件中包含相同的信息,考虑分组存放;在命名中考虑版本信息^[41]。

(9)数据描述与元数据。元数据是描述资源特征的结构化信息,对数据进行详细记录是数据集可发现和可获取的前提。使用元数据目的是为了更好地了解数据,揭示数据的内容特征,不同的学科领域具有不同的元数据标准。如地理空间元数据标准(Content Standard for Digital Geospatial Metadata)、天文学元数据标准(Astronomy Visualization Metadata)、生态学元数据标准(Ecological Metadata Language)等,具体取决于研究项目的数据类型、数据格式和学科领域。如果现有的元数据标准不能满足项目的数据描述需求,研究人员可创建元数据方案,包括描述性版

块、权利版块、结构化版块和技术信息版块;描述性版块即描述数据集内容的信息,如题名、日期、创建者姓名等;权利版块包含数据归属权、数据的使用和访问权限等;结构化版块包括数据的逻辑结构、表现形式等;技术信息版块指数据的技术特征、数据创建软件、使用方法等^[42]。元数据是规范化管理数据的基础,是数据管理计划的重要组成部分,在项目启动前就应确定元数据记录方案。

(10)数据质量控制。数据质量控制对抽样检查数据值的准确性进行二次数据录入,对比检查数据是否有误,分组排序,查找离散值和缺失值、统计计算极端值和异常值。还可以使用 OpenRefine 等数据清洗工具。在数据管理计划中还应包含数据收集的质量控制说明,包括使用的刻度标准、样本二次采集和测量、数据采集标准、数据准入标准、数据验证和使用的受控词汇表等^[43]。

(11)数据处理与分析(技术、工具)。在完成数据的收集和组织的后,根据数据类型和预期结果选取恰当的数据分析方法对数据进行处理和分析是得出研究结论的关键所在。随着数据量的剧增和数据问题的日益复杂化,研究人员在处理和分析数据时往往需要借助专业的数据分析工具,如 SAS、SPSS、STATA 等。此外,在进行数据建模、仿真计算时还需要研究人员掌握一定的计算机编程等技术。

(12)数据解读。数据解读能力主要强调认识数据的内容特征和洞察数据的实质内涵。这要求研究人员具备三方面的能力:一是扎实的专业领域知识;二是敏锐的数据分析能力;三是精准的数据表达能力。任何数据认知错误和表达错误都可能造成研究结论的巨大差异,导致严重后果。

(13)数据可视化。无处不在的图像和可视化媒体正在改变 21 世纪的素养格局,可视化素养被视为公民的必备素养^[44]。大数据环境下,利用数据可视化技术形象、直观地展示数据内容和研究结论已成为学术信息交流的重要方式。Gephi、Exhibit 和 Gnuplot 等数据可视化软件在科学研究中发挥着重要作用,研究人员应学习并掌握相应的数据可视化技能。

(14)数据安全与备份(数据保存环境与存储介质)。采取数据安全措施能够有效避免数据被不当访问、使用、修改、传播和破坏,数据安全的含义主要包括:一是网络安全,确保机密数据不能上网,将敏感信息保存到不能联网的计算机中;二是物理安全,确保重要资料存放地的安全(如保安看守),寻求可信的人充当计算机的故障检修员;三是计算机系统和文件安全,包括更新计算机杀毒软件、避免使用邮件或 FTP 协议传输保密数据,在计算机和数据文件中使用密码等加密措施;另一种确保数据安全的方法是定期进行数据备份。进行可靠地数据备份是数据管理实践的重要环节。定期数据备份能够有效规避数据丢失、数据损坏、病毒或黑客破坏、停电、软件损坏以及其他人工错误。建议对数据进行三份备份,即原始数据、原始数据在本地的备份、原始数据在远程设备上的备份^[45]。

(15)数据版本控制。版本控制是对数据处理不同阶段的数据进行标识的过程。它能够对数据的处理流程进行跟踪,标识哪个是草案,哪个是修订以及哪个是最新的版本。能够有效避免数据的混乱,尤其是研究小组的人员同时对数据进行处理时。总的来说,版本控制的优点包括可追溯性、可识别性、清晰明晰、减少重复、减少错误等。版本控制一般通过对文件命名来体现,此外还可使用版本控制表,即对文件进行操作修改时,填写文件版本的修订情况,包括新版本的序号、修改人、修改目的、修改日期等^[46]。

(16)数据保存(短期保存和长期保存)。数据保存是数据管理流程的关键步骤,不仅指项目结题后的数据存档,还包括项目进程中数据分析与处理等阶段的数据保存。在项目期间需要短期地保存数据时,应注意数据类型与格式、数据存储介质与迁移、数据备份与安全等问题;在制定数据管理计划时应

明确提出项目结题后的数据长期保存方案,具体包括数据存储地点(如公共数据中心、领域数据仓储、机构知识库)、数据保存内容、数据保存时间、数据保存介质、数据保存成本、敏感隐私性数据存储策略等方面的问题。

(17)数据遴选与评估。数据保存面临着高昂的数据保存费用、需要花费更多的劳动和精力、在良莠不齐的海量数据中进行检索和查询使得数据发现变得更加困难等挑战,并不是所有的数据都需要保存。因而,制定数据存储策略时应遴选和评估数据,即确定哪些数据需要存储,哪些数据需要被销毁和删除,评估数据是否具有未来的利用价值。在遴选和评估数据时,可根据以下问题进行判断:资助机构和学习要求保存什么数据?数据对于研究项目和机构而言是否具有重要价值?数据中是否涉及知识产权和相关伦理问题?是否对数据进行充分描述?是否能够支付数据保存费用?

(18)数据发表。目前,科研数据发表主要有三种形式:一是科研数据独立发表,将科研数据集作为独立的数字对象存储在数据仓储或数据中心,支持数据的发现、获取和再利用;二是发表数据论文,在数据论文中描述数据集的内容及其相关信息,将数据论文发表到专业数据期刊中,原始数据集存储在期刊出版商或其指定的数据仓储中;三是数据作为期刊论文的补充材料发表。在论文中注明数据来源及获取方式,根据论文与数据之间的关联获取原始数据,支持研究结论验证,防止学术不端^[47]。其中,将数据作为期刊论文的补充材料发表是最普遍的数据发表形式,分两种情况:第一,论文发表时将相关数据存储到数据仓储中,通过 DOI 等数据唯一标识符与论文创建关联;第二,数据作为论文的补充材料同论文一并提交到期刊出版商。了解数据发表的不同形式,并基于实际需求选择恰当的数据发表方式,是研究人员传播学术成果、获得学术声誉的基本常识。

(19)数据共享与许可协议、数据访问权限。在共享数据时设置数据访问权限、选择适当的数据许可协议是保障数据创建者权益的重要手段。共享的数据涉及商业机密、敏感信息和隐私内容时,可以设置数据库或数据仓储的数据访问权限,通过密码等方

式控制数据的访问,同时保证数据的可发现性和可访问性。数据使用权限主要通过数据许可协议体现。目前使用较广泛的是 Open Data Commons^[48](分为 PDDL、ODC-By、ODC-ODbL 三种级别)和 Creative Commons^[49](分为 BY、NC、ND、SA 四种级别)制定的数据许可协议。创建者根据数据开放需求选择恰当的数据许可协议,数据使用者遵守相应的数据使用规范,能够有效避免数据知识产权纠纷等法律问题,促进数据开放共享环境的健康有序发展。

(20)数据共享的伦理问题。数据共享的伦理问题主要体现在数据的隐私性、敏感性和保密性。社会科学、生物医学等领域的项目往往包含人类或动物的隐私信息,或是能够识别个人信息的敏感数据。凡是涉及人类受试者或濒危动物信息的研究项目需通过伦理审查方可实施。也就是说,在开展调查前,需获得受试者的知情同意;在数据处理和分析时,应采用匿名化处理等方式保护受试者的信息不泄露;在共享数据时,需要获得受试者的数据共享知情同意许可,也可以限制数据共享的范围和对象来保护受试者的隐私信息。此外,涉及商业机密、专利信息、尚未发表的论文数据,以及公开将导致不良后果的敏感数据(如毒品信息)等也是属于数据伦理不宜公开的范畴。

(21)数据共享的法律问题。首先是数据的知识产权和版权问题。英国、澳大利亚等国认为数据知识产权赋予数据创建者处理数据的权利,即允许其他研究者做什么和不能做什么,而中美等国认为数据作为事实材料不受知识产权的保护,而使用仪器设备、花费时间成本和资源成本来收集、处理、分析数据,这种原始的数据集可被视为知识产权保护的范围;其次是数据的归属权问题。一般而言,研究人员在职务岗位或利用机构环境所产生的数据归属于本机构而非数据创建者。此外,数据共享的法律问题还包括合同、协议规定的其他具体问题。

(22)数据引用。随着数据集被视为一种独立的科研成果,数据引用也随之成为衡量学术影响力的重要指标。对于数据生产者而言,数据引用指标与传统文献引用一样,是对数据生产者研究工作的肯定和支持,代表学术贡献和学术声誉;对于数据利用者而言,规范化的数据引用有利于准确查找数据,更好地理解数据和获取数据,进而促进新的知识发现。数据引用的构成要素包括但不限于:数据创建者、数据集名称、出版年、出版商、可用性和访问情况(如 URL 或 DOI)。不同的数据仓储、期刊出版商和组织机构具有不同的数据引用格式,如 DataCite 的数据引用格式为:创建者(发布年):标题.出版商.标识符^[50]。在进行数据引用时,应先了解具体的数据引用标准格式。此外,EndNote 等文献管理软件也添加了数据引用功能,可以直接创建数据集引用格式。

5 结语

数据素养除了强调数据意识的树立和数据知识的了解外,更重要的是掌握一系列可操作的数据技能集合,涉及从数据产生与收集、数据分析与处理、数据发表与共享到数据再利用的科研数据生命周期全过程,而数据生命周期又与科研工作流密不可分,嵌入项目启动、项目实施、项目结题的各个环节。因此,构建围绕科研工作流和数据生命周期的数据素养能力体系对于科研数据管理服务开展和数据素养教育实施具有重要意义。本文在调研国内外数据素养能力体系研究现状、分析数据素养能力框架研究成果的基础上,围绕科研数据生命周期八个阶段构建了包含 22 项能力的数字素养核心能力框架。如有疏漏,敬请业界前辈批评指正。未来将基于此框架调研不同研究群体的数据素养能力现状,设计嵌入研究过程的数据素养教育模式,培养研究群体的数据意识和数据管理能力。

参考文献:

- [1] Koltay T. Data literacy for researchers and data librarians[J]. Journal of Librarianship and Information Science, 2015:096100-0615616450.
- [2] Schield M. Information literacy, statistical literacy and data literacy[J]. IASSIST Quarterly, 2004, 28(2/3):6-11.
- [3] Jian Q, D'Ignazio J. Lessons learned from a two-year experience in science data literacy education[C]. 31st Annual IATUL

- Conference. Purdue University, 2010.
- [4] Calzada Prado J, Marzal M A. Incorporating data literacy into information literacy programs: Core competencies and contents [J]. Libri, 2013, 63(2):123-134.
- [5] 郝媛玲, 沈婷婷. 数据素养及其培养机制的构建与策略思考[J]. 情报理论与实践, 2016, 39(1):58-63.
- [6] Carlson J, Fosmire M, Miller C C, et al. Determining data information literacy needs: A study of students and research faculty [J]. portal: Libraries and the Academy, 2011, 11(2):629-657.
- [7] Ogier A L, Lener E, Miller R K. The data literacy advisory team at virginia tech: Developing a content model for data literacy instruction [EB/OL]. [2016-04-30]. <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1054&context=dilsymposium>.
- [8] 张静波. 大数据时代的数据素养教育[J]. 科学, 2013, 65(4):29-32.
- [9] 肖潇. 基于数据生命周期的科学数据服务模式研究[D]. 北京: 中国科学院研究生院, 2012:16.
- [10] The research lifecycle [EB/OL]. [2016-04-30]. <https://www.jisc.ac.uk/guides/implementing-a-virtual-research-environment-vre>.
- [11] Research data lifecycle [EB/OL]. [2016-04-30]. <http://www.data-archive.ac.uk/create-manage/life-cycle>.
- [12] DCC curation lifecycle model [EB/OL]. [2016-04-30]. <http://www.dcc.ac.uk/resources/curation-lifecycle-model>.
- [13] The data life cycle [EB/OL]. [2016-04-30]. <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>.
- [14] Best practices [EB/OL]. [2016-04-30]. <https://www.dataone.org/best-practices>.
- [15] DDI data lifecycle [EB/OL]. [2016-04-30]. <http://www.ddialliance.org/training/why-use-ddi>.
- [16] Research data management [EB/OL]. [2016-04-30]. <http://guides.library.vcu.edu/data>.
- [17] Humprey C. e-Science and the Life Cycle of Research [EB/OL]. [2016-04-30]. <http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc>.
- [18] Association of College and Research Libraries. Working Group on Intersections of Scholarly Communication and Information Literacy. Intersections of scholarly communication and information literacy: creating strategic collaborations for a changing academic environment [M]. Association of College and Research Libraries, 2013:28.
- [19] Konkiel S, Marshall B, Polley D. Integrating data management literacies with data visualization instruction a one-shot workshop [EB/OL]. [2016-04-30]. <http://docs.lib.purdue.edu/dilsymposium/2013/posters/3/>.
- [20] Tygel A, Kirsch R. Contributions of Paulo Freire for a critical data literacy [EB/OL]. [2016-04-30]. <http://www.dataliteracy.eita.org.br/wp-content/uploads/2015/02/Contributions-of-Paulo-Freire-for-a-critical-data-literacy.pdf>.
- [21] Schneider R. Research data literacy [C]. European Conference on Information Literacy. Springer International Publishing, 2013:134-140.
- [22] Ridsdale C, Rothwell J, Smit M, et al. Strategies and Best Practices for Data Literacy Education [EB/OL]. [2016-04-30]. http://www.mikesmit.com/wp-content/papercite-data/pdf/data_literacy.pdf.
- [23] 隆茜. 数据素养能力指标体系构建及高校师生数据素养能力现状调查与分析[J]. 图书馆, 2015(12):51-56, 62.
- [24] What is research data? [EB/OL]. [2016-04-30]. <http://guides.library.yale.edu/content.php?pid=324929&sid=2665407>.
- [25] Data management [EB/OL]. [2016-04-30]. <http://libraries.mit.edu/data-management/>.
- [26] Dissemination and Sharing of Research Results [EB/OL]. [2016-04-30]. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
- [27] NIH Data Sharing Policy [EB/OL]. [2016-04-30]. http://grants.nih.gov/grants/policy/data_sharing/.
- [28] NASA Data & Information Policy [EB/OL]. [2016-04-30]. <http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/>.
- [29] RCUK Common Principles on Data Policy [EB/OL]. [2016-04-30]. <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>.
- [30] OECD Principles and Guidelines for Access to Research Data from Public Funding [EB/OL]. [2016-04-30]. <http://www.oecd.org/sti/sci-tech/38500813.pdf>.
- [31] Funding Agency and Data Management Guidelines [EB/OL]. [2016-04-30]. <https://www.lib.umn.edu/datamanagement/funding>.
- [32] 科学数据共享工程 [EB/OL]. [2016-04-30]. <http://www.sciencedata.cn/index.php>.

- [33] 国家科技计划项目科学数据汇交暂行办法(草案)[EB/OL].[2016-04-30].<http://www.sciencedata.cn/fagui.php>.
- [34] 现代图书情报技术.支撑数据提交要求[EB/OL].[2016-04-30].http://manu44.magtech.com.cn/Jwk_infotech_wk3/fileup/1003-3513/NEWS/20160408165409.pdf.
- [35] BBSRC data sharing policy[EB/OL].[2016-04-30].<http://www.bsrc.ac.uk/web/FILES/Policies/data-sharing-policy.pdf>.
- [36] DMP Tool[EB/OL].[2016-04-30].<https://dmp.cdlib.org/>.
- [37] DMP Online[EB/OL].[2016-04-30].<http://www.dcc.ac.uk/dmponline>.
- [38] DMP Assistant[EB/OL].[2016-04-30].<https://assistant.portagenetwork.ca/>.
- [39] Data types & file formats[EB/OL].[2016-04-30].<http://data.library.virginia.edu/data-management/plan/format-types/>.
- [40] Managing research data[EB/OL].[2016-04-30].<https://data.bris.ac.uk/>.
- [41] Data management support for researchers[EB/OL].[2016-04-30].<http://www.gla.ac.uk/services/datamanagement/>.
- [42] Research data management[EB/OL].[2016-04-30].<http://library.uoregon.edu/datamanagement>.
- [43] Strategies for sharing your data[EB/OL].[2016-04-30].<http://data.research.cornell.edu/>.
- [44] A practical guide to ACRL's Visual Literacy Competency Standards [EB/OL].[2016-04-30].<http://www.ala.org/news/member-news/2016/03/practical-guide-acrl-s-visual-literacy-competency-standards>.
- [45] Research data[EB/OL].[2016-04-30].<http://www.bath.ac.uk/research/data/>.
- [46] Research data Management[EB/OL].[2016-04-30].<http://www.data.cam.ac.uk/>.
- [47] 陈秀娟,吴鸣,胡卉.嵌入科研工作流的图书馆数据管理服务——以化学学科为例[J].图书馆论坛,2016,36(3):49-55,102.
- [48] Open Data Commons[EB/OL].[2016-04-30].<http://opendatacommons.org/licenses/>.
- [49] Creative Commons[EB/OL].[2016-04-30].<https://creativecommons.org/>.
- [50] Datacite[EB/OL].[2016-04-30].<https://www.datacite.org/>.

作者简介 胡卉,女,中国科学院文献情报中心、中国科学院大学硕士研究生;吴鸣,女,中国科学院文献情报中心研究员。

(上接第95页)

- [12] Hamasaki M,Goto M,Takeda H.Social Infobox:collaborative knowledge construction by social property tagging [C]. Proceedings of the ACM 2011 conference on Computer supported cooperative work.ACM,2011.
- [13] Almoqhim F,Millard D,Shadbolt N.An Approach to Building High-Quality Tag Hierarchies from Crowdsourced Taxonomic Tag Pairs[M].JATOWT A,LIM E-P,DING Y,et al.Social Informatics.Springer International Publishing,2013:129-138.
- [14] 王爽,徐行.基于用户分类标签建立结构性的大众分类法[J].图书馆学研究,2011(5):73-76.
- [15] 尚加宁,韩露盈.图书分类法性能的定量测评初探[J].情报理论与实践,1999,22(5):356-358.
- [16] Solskinnsbakk G,Gulla J A,Haderlein V,et al.Quality of hierarchies in ontologies and folksonomies[J].Data & Knowledge Engineering,2012(74):13-25.
- [17] Golder S A,Huberman B A.Usage patterns of collaborative tagging systems [J].Journal of Information Science,2006,32(2):198-208.
- [18] Lin C S,Chen Y F.Examining social tagging behaviour and the construction of an online folksonomy from the perspectives of cultural capital and social capital[J].Journal of Information Science,2012,38(6):540-57.

作者简介 张鹏翼(1981-)女,北京大学信息管理系副教授;苏杰(1990-)男,北京大学信息管理系硕士研究生;张久珍(1974-)女,北京大学信息管理系副教授。