

# 生命科学领域研究生数据素养现状与需求调查分析

——以中国科学院大学为例

胡卉<sup>1</sup>, 吴鸣<sup>1,2</sup>, 欧阳峥峥<sup>3</sup>

(1. 中国科学院文献情报中心; 2. 中国科学院大学经济与管理学院; 3. 中国科学院成都文献情报中心)

**摘要:** 本研究基于国内外科研数据管理服务和数据素养教育课程的经验, 以生命科学领域作为切入点, 以中国科学院大学研究生为调查对象, 设计了面向生命科学领域的研究生数据素养现状与需求调查问卷, 从科研数据概况、制定数据管理计划、数据收集与记录、数据处理与分析、数据管理与保存、数据发表与共享、数据权益与伦理等角度揭示研究生的数据素养现状和需求特点, 以期为更好地完善数据素养教育和科研数据管理服务提供参考和依据。

**关键词:** 数据素养; 生命科学; 中国科学院大学

中图分类号: G254.97; G258.6 文献标志码: A 文章编号: 1005-8214(2018)06-0050-07

DOI:10.14064/j.cnki.issn1005-8214.2018.06.011

## Survey and Analysis of Data Literacy Status and Needs of Postgraduate in Life Science

—— A Case Study of University of Chinese Academy of Sciences

Hu Hui, Wu Ming, Ouyang Zheng-zheng

**Abstract:** This article, based on experience in scientific research data management services and data literacy education courses at home and abroad, designs questionnaire to investigate the status and demand of data literacy of the postgraduate in the field of life science in University of Chinese Academy of Sciences. The questionnaire reveals the status and demand of data literacy from several aspects, which can provide reference for better data literacy education and scientific research data management service.

**Keywords:** Data Literacy; Life Science; University of Chinese Academy of Sciences

数据是开展科学研究的根基所在, 培养研究人员的数据意识、开展数据素养技能培训的重要性不言而喻。研究生是未来科研事业的参与者, 在知识学习和专业实践的早期阶段培养其数据管理技能和数据素养能力, 有利于他们更好地适应学习生活和进行科研实践。因此, 本研究选取中国科学院大学(以下简称中科院)研究生作为调查对象, 设计了“中国科学院大学研究生数据素养现状与需求调研”问卷, 以期在了解中科院研究生的数据素养现状和需求基础上, 为更好地完善数据素养教育和科研数据管理服务提供参考和依据。

### 1 调查问卷设计

调查问卷设计的前期工作主要包括三个方面: ① 调研和总结国内外科研数据管理服务和数据素养教育课程经验, 如, 分析 DCC<sup>[1]</sup>、DataONE<sup>[2]</sup>、马萨诸塞大学医学院图书馆、爱丁堡大学图书馆等机构的数据

素养课程框架和数据管理服务内容体系, 提取关键主题和重要概念;<sup>[3]</sup> ② 参考和借鉴国内外学者编制的研究生数据素养调查表, 结合本研究的调查目的和内容, 搭建调查问卷的主体框架; ③ 以生命科学领域作为切入点, 分析和归纳生命科学领域科研数据的特点(包括数据类型与格式、所在领域的数据仓储、常用数据处理工具等), 并以此为基础, 设计了本研究的调查问卷, 调查问卷包括两个模块。

(1) 调研背景和调查目的。问卷的开场白强调了在数据驱动的新型科研环境下对科研数据进行规范管理的重要性, 以及科研实践中科学工作者面临的数据管理困境, 指出本次调研旨在为图书馆更好地设计数据管理课程和开展培训服务提供参考和依据。

(2) 问卷主体。分为调研对象的基本信息和调研的主要内容两大部分, 共计 30 道题目, 其中客观题 28 道, 开放性题目 2 道。① 调查对象的基本信息。

包括调查对象所在研究机构、专业、具体研究方向三个问题，要求调查对象根据自身情况如实填写，确保调查对象和调查内容相匹配。② 调研的基本内容包括数据素养现状和需求两方面（见表 1）。现状调研以数据素养内涵为基础，涵盖数据意识、数据知识、数据技能及数据伦理道德四个层面的内容，结合科研数据生命周期流程，构建了包括科研数据概况、制定数据管理计划、数据收集与记录、数据处理与分析、数据管理与保存、数据发表与共享、数据权益与伦理七个模块的现状调查框架。在调研数据素养现状的基础上，对研究生接受数据素养教育的态度、内容和方式进行考察，并以开放式问题的形式询问调研对象在科研数据管理过程中遇到的问题和障碍，以及对数据管理和数据素养课程的建议。

表 1 调查问卷内容结构

调查问卷基本框架		调查具体内容	题号
调查对象的基本信息		调查对象所在研究机构、所学专业、研究方向	1-3
数据素养现状	科研数据概况	为什么要管理科研数据；什么是科研数据；对科研数据管理与共享的相关政策和要求的认知情况；对本学科科研数据类型、数据格式、数据量的认知情况	1-6
	制定数据管理计划	对科研数据生命周期、数据管理计划以及制定数据管理计划工具的认知情况	7-9
	数据收集与记录	记录科研数据的方式	10
	数据处理与分析	处理科研数据的软件和工具 所用科研数据的来源 科研数据质量控制方法	11-13
	数据管理与保存	科研数据的描述和说明情况(元数据)	14
	数据发表与共享	对数据共享与发表途径的认知情况 适用的科研数据共享激励措施	15-16
	数据权益与伦理	数据引用现状 对科研数据权益的认知情况	17-18
数据素养需求	数据素养教育需求态度	对研究生在从事科研工作前接受数据管理和数据素养教育的看法	19
	数据素养教育需求内容	针对科研数据管理生命周期流程各阶段所需的数据素养知识和技能	20-24
	数据素养教育需求方式	希望通过哪些方式来了解科研数据管理的内容	25
	开放性问题	在科研数据管理过程中遇到的问题和障碍；对数据管理和数据素养课程的建议	26-27

## 2 调查实施

本次问卷的调查对象为中国科学院大学生命科学领域研一新生，调查对象的选取主要考虑以下两个方面。

(1) 生命科学领域的学科特点。生命科学作为中科院的重点学科，每年培养包括中国科学院大学生命科学学院在内，以及植物所、动物所、遗传与发育研究所、微生物研究所等研究生委培单位在内的上千名研究生。而生命科学领域科研数据的来源和形式多样，既包括基因测序，又包括分子通道等不同类型模式的科研数据，以及不同人群受试数据等。因此，本

次调查以生命科学领域作为切入点。

(2) 研一新生接受数据素养教育的必要性和可行性。研究生是未来科研事业的主力军，而中国科学院大学研究生培养为两段式模式，即研一的基础课程集中教学阶段和研二后回到研究所进行科研实践。在研一集中教学期间，更多的是专业基础知识和基本研究技能的学习和培养。因此，通过学分课程、培训讲座等方式预先培养学生在未来的科研实践工作中所需的科研数据管理能力具有可行性。

正式的问卷调查开始于 2016 年 9 月下旬，主要借助问卷星平台开展调研，通过课程网站发送问卷链接，并在课上集中填答。本次调查共发放问卷 60 份，回收问卷 59 份，回收率 98.3%，其中有效问卷 59 份，问卷有效率 100%。

## 3 调查结果分析

问卷回收后采用 Excel 进行数据统计和分析，通过频数、百分比描述性统计等指标，探索问卷结果中所反映的中科院研究生的数据素养现状与需求规律。

### 3.1 调查对象基本概况

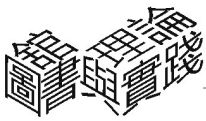
从调查对象所在机构来看，这 59 名研一学生主要分布于中国科学院不同的科研教育单位，分别是植物研究所、西双版纳热带植物园等 22 所机构。从调查对象所学专业来看，涵盖了遗传学、生物化工等 18 个学科专业和多个研究方向；调查样本的选取符合中科院生命科学领域研究分布实况，具有良好的代表性。

### 3.2 数据素养现状分析

#### 3.2.1 科研数据概况

题项 1-6 主要考察研究生对科研数据及数据管理基本知识的认知情况。

(1) 科研数据管理的重要性。问卷中该题目 1 表述为“您认为科研数据管理给科研带来哪些益处”。问题的选项设置参照数据管理国际组织 UK Data Archive 对数据管理意义的表述，<sup>[4]</sup>通过调查研究生对“科研数据管理的意义”的认识，揭示中科院研究生对科研数据管理的认知程度和积极态度。调查结果显示，接受调查的大多数研究生能较全面地认识到科研数据管理的意义，他们认为管理科研数据的主要意义在于数据再利用，有利于促进科学创新(89.83%)和减少数据重复收集的成本(86.44%)，能使研究发现和学术成果可检验(83.05%)，其次，可促进科研更加透明公开(79.66%)、引导数据生产者和数据利用者加强新合作(77.97%)，并促进研究方法改进和验证(76.27%)。综上，接受调查的研究生已经认识到科研数据管理的



重要性,并且能够清晰地辨析科研数据管理对其研究工作和学术声誉各方面的意义所在。

(2) 对科研数据的认知情况。问卷中该题目 2 表述为“您所了解的科研数据包括哪些”。问题的选项设置参照国际数据管理专业组织数据管理培训课程中对“广义新型科研数据”的介绍。<sup>[5]</sup>通过考察研究生对科研数据的认知情况,揭示中科院研究生对科研数据这一概念的了解程度。调查结果显示,接受调查的研究生对所在学科的科研数据具有良好的认知度,但对采访记录、文本和动画材料等广泛意义上的科研数据缺乏了解。其中,认知度最高的仍是统计和测量数据(88.14%)、实验及模拟数据(86.44%)等传统型科研数据,其次,是从相机或实验设备中导出的图片(76.27%)、观测数据(71.19%)、模型和软件(55.93%)。但就文本和动画材料、采访记录等广义科研数据而言,仅有 1/3 的学生将其视为科研数据。此外,有 3.39% 的学生表示不清楚什么是科研数据。

(3) 对科研数据管理与共享的相关政策和要求的认知情况。问卷中该题目 3 表述为“您了解哪些科研数据管理与共享的相关政策和要求”,问题的选项设置参考顾立平教授提出的科学数据开放获取政策研究框架。<sup>[6]</sup>该题主要考察研究生对科研数据管理与共享政策的了解情况。调查结果显示,接受调查的研究生对数据管理与共享政策的总体了解程度较低。其中,研究生了解的主要科研数据管理政策包括所在实验室或研究组的数据管理条例和规则政策(50.85%),其次,是所在机构的数据管理政策和要求(33.9%)、期刊出版商的数据提交和存缴要求(25.42%)、数据仓储的数据政策和要求(22.03%)。对国家层面、科研资助机构以及国际组织的数据管理与共享政策知之甚少。此外,超过 1/4 的学生表示完全不了解科研数据管理与共享的政策和要求。

(4) 对本学科科研数据类型的认知情况。问卷中该题目 4 表述为“您所在学科领域的科研数据包括哪些类型”,题目选项设置参照弗吉尼亚理工大学的科研数据分类体系。<sup>[7]</sup>通过考察学生是否了解本学科领域的科研数据类型,以及本学科领域科研数据的具体所属类型,由此可以较好地反映学生对本领域科研数据的熟悉程度。调查结果显示,98.31% 的研究生表示了解并能明确界定本学科科研数据所属的类型,其中,最主要的数据类型是实验型数据(98.31%),其次,是引用或规范数据(64.41%)和观测型数据(64.41%),

而仿真数据(16.95%)、派生或编译数据(13.51%)所占比例较少。

(5) 对本学科科研数据格式的认知情况。问卷中该题目 5 表述为“您所在学科领域的科研数据包括哪些格式”,题目选项设置主要参考国外图书馆科研数据管理服务和教学课程中对不同数据格式的介绍。通过考察学生是否了解本学科领域的科研数据格式,以及本学科领域科研数据的具体格式,可以较好地反映学生对本领域科研数据的熟悉程度。调查结果显示,98.31% 接受调查的研究生表示了解并能明确界定本学科科研数据的格式,其中,最普遍的数据格式是表格(91.53%),其次,是图像(谱图)(89.83%)和文本(89.83%)、特定实验仪器采集的数据(69.49%)、模型(44.07%)以及软件语言(程序代码)(35.39%),而视频和音频格式的科研数据则所占比例较小。

(6) 对本学科科研数据量的认知情况。问卷中该题目 6 表述为“您所在学科领域的科研数据量大约是什么级别”,通过考察学生是否了解本学科领域的科研数据量,以及本学科领域科研数据的具体量级,可以较好地反映学生对本领域科研数据的熟悉程度。调查结果显示,接受调查的研究生对所在学科领域的科研数据量认知度较低,只有 54.24% 的研究生表示了解并能界定本学科领域的科研数据量,另外 45.76% 的学生表示不了解或不确定本学科领域的科研数据量级。

### 3.2.2 数据管理计划

选项 7-9 主要考察学生在制定数据管理计划阶段的数据素养现状。

(1) 对科研数据生命周期的认知情况。问卷中该题目 7 表述为“您了解科研数据生命周期吗”。通过考察学生对科研数据生命周期的认知情况,揭示学生对整个科研流程和数据管理流程的认识与掌握。调查结果显示,接受调查的研究生对科研数据生命周期的认知度较低,其中,只有 20.34% 的学生表示“比较了解”,而 79.66% 的学生表示不了解或不确定什么是科研数据生命周期。

(2) 对科研数据管理计划的认知情况。问卷中该题目 8 表述为“您认为一份数据管理计划应包含哪些内容”,题目选项设置主要参考美国国家科学基金委发布的数据管理计划模板,将其中四大要素设置为本题目的选项。<sup>[8]</sup>通过考察学生是否了解数据管理计划,以及数据管理计划应包含的具体内容,可以较好地体现学生对数据管理计划的认知程度和掌握情况。调查结果显示,接受调查的 90% 的研究生表示了解并



能明确数据管理计划所包含的内容。数据管理计划四项基本要素中,接受调查的研究生认为最重要的是数据存储和数据共享策略(79.66%),其次,是科研项目中预计产生的数据和资料(72.88%),以及数据再利用、再传播、数据衍生产品的规定和限制(64.41%),拟采用的数据管理标准和方法(57.63%)。另有10.17%的学生表示不了解科研数据管理计划及其包含的内容。

(3) 对制定科研数据管理计划工具的认知情况。问卷中该题目9表述为“您了解制定数据管理计划的工具吗”,题目选项设置主要列举了目前国际通用的几种数据管理计划工具。通过调研学生对科研数据管理计划工具的认知情况,揭示学生对数据管理计划的了解程度,以及他们制定和撰写数据管理计划的能力水平。调查结果显示,接受调查的研究生对数据管理计划工具的认知度低,其中89.83%的学生不了解数据管理计划工具,也不知道如何利用工具创建数据管理计划。

### 3.2.3 数据收集与记录

题项10即记录科研数据的方式。旨在考察学生在数据收集与记录阶段的数据素养现状,问卷中该题目表述为“您了解并使用的本学科记录科研数据的方式有哪些”。调查结果显示,大多数接受调查的研究生了解并能够较好地掌握本学科领域科研数据的记录方法(98.31%)。在对科研数据进行记录时,使用最广泛的科研数据记录方式是U盘或移动硬盘等便携式存储设备(94.92%)、个人电脑(94.92%)、纸质笔记本或实验室记录本(93.22%)以及实验室或办公室电脑(86.44%),其次,是网盘或云盘等网上存储空间(57.63%),此外,还有学生利用服务器来记录和保存科研数据。

### 3.2.4 数据处理与分析

题项11-13旨在考察学生在数据处理与分析阶段的数据素养现状。

(1) 对科研数据处理所需工具和软件的了解和掌握情况。问卷中该题目11表述为“本学科在处理科研数据时,需要使用哪些软件和工具”,选项中罗列了生命科学领域常用的数据处理工具,考察学生对本学科领域科研数据处理工具的了解程度,以及对这些数据处理工具的掌握水平。调查结果显示,大多数接受调查的研究生表示了解本学科领域的科研数据处理工具(89.83%)。具体而言,生命科学领域数据处理工具以基本的数据处理工具为主(如Excel占比93.22%,

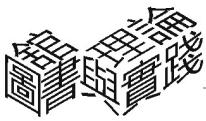
统计分析软件SPSS占比72.88%),重视多学科数据分析工具的融合与利用(如用于统计分析与绘图的R语言环境占比40.68%,商业数学软件Matlab占比27.12%),同时,由于生命科学领域的学科专业性,在数据处理和分析过程中还使用较多的专业性软件,如Primer、Origin Demo、GraphPad PRISM等。此外,10.17%的学生表示对科研数据分析和处理的工具并不了解。

(2) 所用科研数据的来源情况。问卷中该题目12表述为“在研究过程中,您使用的科研数据来源包括哪些”。本题旨在考察学生在开放科学开放数据、数据再利用的科研大背景下,用于研究实践的数据来源,对学生的自主实验数据、团队实验数据,以及从外部获取的科研数据进行辨识。调查结果显示,部分接受调查的研究生具有广泛的科研数据来源,既有自己和团队的实验数据,还包括从文献、数据中心、数据仓储下载的数据(59.32%);但仍有部分学生的科研数据来源渠道较窄,有27.12%的学生表示在研究过程中只使用自己及所在团队的实验数据,10.17%的学生表示研究过程中全部是自己的实验数据。此外,有3.39%的学生表示对自己研究过程中所涉及数据的来源尚不明确。

(3) 科研数据质量控制方法。问卷中该题目13表述为“您认为控制和评估科研数据质量的方法有哪些”。问题选项设置参考张晓林教授对科学研究方法的表述,设置了数据真实性、数据完整性、数据规范性、数据重现性、数据重复性等科研数据质量评价选项。<sup>[9]</sup>调查结果显示,接受调查的研究生在科研数据质量控制和评价方面表现出较高的认知度,其中,最重要的指标是数据真实性(98.31%),其次,是数据完整性(91.53%)、数据规范性(88.14%),以及数据重现性(83.05%)和数据重复性(81.36%)。

### 3.2.5 数据管理与保存

科研数据的描述与说明情况(元数据)。问卷中该题目14表述为“您认为科研数据应如何进行描述和说明(元数据)”。实现数据可理解、可重用的前提是使用元数据对科研数据进行规范描述和说明,因此,本题目主要考察学生对于数据描述和说明的态度及现状。调查结果显示,接受调查的研究生在数据描述和说明方面具有较高的认知度,绝大多数学生认为应遵循研究团队或实验室统一规范的数据描述和说明规则(94.92%),也可以通过自己设定规则来描述数据(33.90%)。而5.08%的学生认为不用对数据进行描述



和说明,此外,还有少数学生不确定如何描述和说明科研数据。

### 3.2.6 数据发表与共享

题项 15-16 旨在考察学生在数据发表与共享阶段的数据素养现状。

(1) 对科研数据共享与发表途径的认知情况。问卷中该题目 15 表述为“您知道科研数据发表的途径吗”,本题目主要考察学生对三种数据发表方式的认知情况。调查结果显示,大多数接受调查的研究生已经知晓科研数据发表的方式(84.75%),就具体的发表途径而言,发表论文时按照期刊要求将数据作为补充信息提交是学生认知度最高的数据发表方式(76.27%),这也是目前国际上最主流的数据发表方式;而对于将数据提交到本学科领域的数据仓储,在数据期刊发表数据论文两种较新的数据发表方式,也有半数以上的学生表示了解。但仍有 15.25% 的学生表示不清楚科研数据发表的途径。

(2) 适用的科研数据共享激励措施。问卷中该题目 16 表述为“通过下列哪些措施,您愿意开放和共享科研数据”。本题目旨在考察学生的数据共享意愿,以及有效的数据开放共享激励措施。调查结果显示,接受调查 94.92% 的研究生具有潜在的数据共享意愿,只有 5.08% 的学生不确定是否共享科研数据。就各种适用的科研数据共享激励措施来看,学生最关注的是通过共享科研数据来提高个人知名度和学术影响力(64.41%),其次,是资金鼓励共享科研数据(59.32%)和优先发表论文(57.63%),对以强制性要求共享科研数据这一措施的认可度较低(40.68%)。

### 3.2.7 数据权益与伦理

题项 17-18 旨在考察学生的数据权益与伦理现状。

(1) 数据引用现状。问卷中该题目 17 表述为“在参考或引用他人数据时,您会采取以下哪些措施”。本题目旨在考察学生的数据伦理道德,即数据引用情况。调查结果显示,绝大多数接受调查的研究生在参考和引用他人数据时都会有相应的数据引用行为,其中,以引用规范标注数据来源的占 96.61%,致谢数据来源作者的占 52.54%。此外,有 5.08% 的学生在参考和引用他人数据时不标注数据来源,1.69% 的学生表示不知道是否/如何进行数据引用。

(2) 对科研数据相关权益的认知情况。问卷中该题目 18 表述为“您了解科研数据管理和利用过程中涉及的相关权益吗”。本题目旨在考察学生对科研数据相关权益的认知情况,选项设置参照顾立平教授提

出的科研数据权益研究框架。<sup>[10]</sup> 调查结果显示,接受调查的研究生对科研数据相关权益具有较高的认知度(86.44%),具体而言,数据归属权占比 79.66%,知情同意权占比 76.27%,其次,是合法商业利益占比 49.15%,国家安全机密占比 47.46%,隐私权占比 35.59%。但仍有 13.56% 的学生表示不清楚科研数据管理和利用过程中涉及的相关权益。

### 3.3 数据素养需求分析

(1) 研究生接受数据素养教育的重要性。在调研中科院研究生的数据素养教育需求之前,首先,调查中科院研究生在从事科研工作前是否有必要接受数据素养教育这一问题的看法,这是后期设计课程和实施教学的基本前提。从调查结果可知,接受调查的研究生对数据管理和数据素养教育具有强烈需求,接受调查的学生均认为研究生在从事科研工作前有必要接受数据素养的相关培训(其中非常重要占比 81.36%,重要占比 18.64%)。由此可见,面向中科院研究生开设数据管理培训和数据素养课程具有必要性。在明确了研究生具有的数据素养教育需求后,重点调研了学生在科研数据生命周期各阶段的数据素养内容需求。题项 20-24 旨在考察学生对数据素养教育的内容需求。

(2) 制定数据管理计划阶段的需求内容。问卷中该题目 20 表述为“关于数据管理计划,您想了解哪些内容”,主要调研学生在数据管理计划阶段所需的知识和技能。调查结果显示,接受调查的研究生对数据管理计划阶段的数据素养需求从高到低分别是数据管理计划工具(89.83%),数据管理、共享的政策和要求(81.36%),科研数据生命周期(71.19%),数据管理计划构成要素(61.02%)。

(3) 数据收集与记录阶段的需求内容。问卷中该题目 21 表述为“关于数据收集与记录,您想了解哪些内容”,主要调查研究生在收集与记录数据阶段所需的知识和技能。调查结果显示,接受调查的研究生对收集和记录数据阶段的数据素养需求从高到低分别是本学科科研数据的类型、格式和数据量(94.92%),科研数据质量控制与评估(83.05%),本学科科研数据的记录方式(81.36%)。

(4) 数据处理与分析阶段的需求内容。问卷中该题目 22 表述为“关于数据处理与分析,您想了解哪些内容”,主要调查研究生在处理和数据分析阶段所需的知识和技能。调查结果显示,接受调查的研究生对数据处理与分析阶段的数据素养需求主要在数据处

理和分析工具 (98.31%) 以及数据可视化 (76.27%) 方面。

(5) 数据管理与保存阶段的需求内容。问卷中该题目 23 表述为“关于数据管理与保存,您想了解哪些内容”,主要调查研究生在管理和保存数据阶段所需的知识和技能。调查结果显示,接受调查的研究生对数据管理和保存阶段的数据素养需求从高到低分别是科研数据保护与安全措施 (91.53%),科研数据保存与备份 (88.14%),本学科领域对科研数据进行描述的元数据标准 (81.36%),科研数据命名规则 (62.71%) 以及科研数据版本控制 (47.46%)。

(6) 数据发表与共享阶段的需求内容。问卷中该题目 24 表述为“关于数据共享与再利用,您想了解哪些内容”,主要调查研究生在数据共享和再利用阶段所需的知识和技能。调查结果显示,接受调查的研究生对数据共享和再利用阶段的数据素养需求从高到低分别是检索和获取本学科科研数据的来源 (98.31%),科研数据发表和共享的途径 (88.14%),科研数据引用规范 (79.66%),科研数据相关的权益问题和保护措施 (77.97%),以及科研数据共享与许可协议 (69.49%)。

(7) 提升数据素养能力的方式。在界定研究生的数据素养内容需求后,调研了学生可接受的数据管理和数据素养教育方式。问卷中该题目 25 表述为“您希望通过哪些方式来了解科研数据管理的内容”,罗列了目前国际上高校图书馆常用的数据素养教育方式。调查结果显示,建立数据管理平台是学生最期望的数据素养教育方式 (71.19%),其次,是培训讲座 (69.49%),微信公众号、博客等自媒体 (67.80%),在线课程 (66.10%),学分课程 (61.02%),而国外高校图书馆使用最广泛的研讨会却在中科院研究生中遇冷,仅占比 27.12%。

(8) 为了更深入、具体地了解研究生的数据素养现状,在以结构化方式调研了科研数据生命周期各阶段学生的数据素养现状后,在问卷的最后设计了开放性问题来调研学生在数据管理过程中面临的难题和挑战、收集学生对数据管理课程的意见和建议,旨在进一步发掘和揭示学生的数据素养需求。

就数据管理过程中面临的难题和挑战方面,调查结果显示,在具体的科研数据研究实践中,接受调查的研究生主要面临以下五个方面的问题:① 科研数据管理意识淡薄、知识薄弱,数据管理的主动性和积极性低;② 数据收集和查找困难,数据检索效率低;③ 庞大的数据量筛选和取舍困难,难以科学地评估

科研数据的质量;④ 数据分析与处理的规范流程不明确,工具软件的选取和使用不清晰,难以实现和解读数据可视化;⑤ 数据保存和管理等。

就研究生对数据管理课程的意见和建议而言,统计结果显示,接受调查的研究生对数据素养课程的建议主要包括以下三个方面:① 结合领域科研数据特点,介绍不同类型数据的来源渠道,以及数据发现与检索的策略方法,提高学生检索与获取数据的能力;② 以实际应用场景为基础,强调数据分析与处理工具的应用,加大课堂教学中的实用性技能比重;③ 课程讲授与课后练习相结合,重视课后练习与交流总结,提高学生的动手实践能力。

#### 4 数据素养现状与需求总结

根据问卷调查结果分析,接受调查的研究生对数据素养教育具有强烈需求。为此,在问卷分析统计的基础上,本研究从科研数据生命周期流程出发,总结了生命学科领域研究生在数据生命周期各个阶段的数据素养需求问题,并提出相应的对策和建议(见表 2)。根据问卷的统计分析结果,可将接受调查的中科院生命学科领域研究生的数据素养现状特点总结如下。

(1) 能够较好地认识到科研数据管理的重要性,并具有强烈的数据素养教育需求。接受调查的大多数研究生能够较全面地认识到科研数据管理的重要意义,包括数据再利用促进科学创新、减少数据重复收集的成本、使研究发现和学术成果可检验、促进科研更加透明公开、引导数据生产者和数据利用者加强新合作、促进研究方法的改进和验证。同时,研究生具有强烈的数据素养教育需求,认为在知识学习和专业实践的早期阶段通过学分课程、培训讲座等方式培养其数据素养技能非常必要。

(2) 科研数据管理知识基础薄弱。调查结果显示,尽管学生能够认识到数据管理的重要性,但对数据管理的相关知识尤其是对学科领域内科研数据管理的专业知识了解较少。具体表现为对模型软件等新型科研数据的辨识度低,对数据管理与共享政策的认知度低,对数据量级等学科领域科研数据的认识模糊。

(3) 科研数据管理专业技能欠缺。研究生以往的数据管理实践往往是个性化的、缺乏规范性,因此在应对整个科研生命周期以及生命周期各个阶段的数据管理难题和挑战时,缺乏系统性的数据管理专业技能,如,数据获取来源渠道单一,数据处理工具选取困难,缺乏数据管理与保存规范,数据保护与安全措施欠缺等。



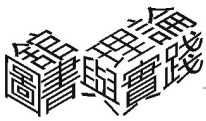


表2 数据素养问题总结与对策建议 (生命学科领域)

不同阶段	问题总结	对策建议
科研数据概况	(1) 对模型软件等新型科研数据的辨识度低 (仅 30% 的研究生能够界定新型科研数据) (2) 对数据管理与共享政策的认知度低 (超过 25% 研究生完全不了解, 其他学生也了解的不全面) (3) 对学科领域内的科研数据了解有待提升 (超过 50% 的研究生不明确本学科数据量级)	(1) 强化数据意识, 提升科研数据管理基础知识水平 (2) 嵌入学科, 培养领域内科研数据管理的专业知识 (3) 加强对数据管理与共享政策的学习, 如科研资助机构、期刊出版商等
数据管理计划阶段	(1) 缺乏对科研数据生命周期和数据管理流程的认识 (80% 的研究生不了解生命周期) (2) 对数据管理计划的重要性及其构成要素的理解不够全面 (超过 10% 的研究生不了解数据管理计划, 其他学生了解也不全面) (3) 对数据管理计划工具的认知度和利用水平低 (90% 的研究生不了解数据管理计划工具)	(1) 加强对科研数据生命周期、数据管理计划等基本概念的理解 (2) 培养利用数据管理计划工具创建符合需求的数据管理计划的能力, 包括制定数据存储与共享的策略、选取合理的数据管理标准和方法等
数据收集与记录阶段	(1) 数据来源渠道单一, 主要局限于自身或所在团队的实验数据 (40% 的研究生), 缺乏对数据仓储等外在资源的发现与再利用	(1) 深入学科, 提升符合学科特性的科研数据采集和获取能力 (2) 加强对数据平台、数据仓储、数据中心等数据来源的学习, 提升检索和获取外部数据的能力 (3) 以科研数据的特点和需求为基础, 选择科学合理的数据记录方式
数据处理与分析阶段	(1) 不能选择恰当的数据处理工具和软件 (10% 的研究生不了解本领域数据处理工具) (2) 对数据处理工具尤其是专业性数据分析工具的使用水平有待加强 (3) 数据质量评估方法不知如何运用于数据质量控制实践 (4) 数据分析结果 (图表) 的解读与表述困难	(1) 根据科研数据的不同类型和格式, 选取恰当的数据处理和分析工具 (2) 加强对数据处理工具和软件的学习 (包括基本的数据处理软件和专业性数据处理软件) (3) 学习科研数据质量控制的方法和措施 (4) 以具体工具为例, 学习分析结果的基本原理和代表含义, 严谨认真地表述分析结果, 得出科学合理的研究结论
数据管理与保存阶段	(1) 缺乏科学的数据管理与保存规范 (2) 数据管理操作能力有待提升 (包括元数据利用、数据备份、数据安全等措施)	(1) 选择恰当的科研数据保护与安全措施 (2) 学习科研数据保存与备份策略 (3) 学习本学科领域的元数据标准, 对科研数据集进行规范描述和说明 (4) 学习科学的数据命名规则 (5) 学习科研数据版本控制措施
数据发表与共享阶段	(1) 缺乏数据共享意识 (即使采取激励措施, 其共享数据的意愿也不明显) (2) 对数据发表的概念、重要性以及方式缺乏认识 (15% 的研究生完全不了解数据发表, 其他研究生也了解得不全面)	(1) 提升数据开放、数据共享意识, 学习共享数据的重要性和必要性 (2) 学习数据发表的方法, 选择恰当的数据发表途径
数据权益与伦理	(1) 数据引用缺乏规范性 (2) 对科研数据相关的权益缺乏全面了解 (15% 的研究生表示完全不了解, 其他研究生也了解得不全面)	(1) 学习数据引用知识, 提升数据引用的规范性和标准性 (2) 学习并遵守科研数据管理涉及的权益问题, 包括数据归属权、知情同意权、合法商业利益、国家安全机密、隐私权等

(4) 具有良好的数据伦理与道德意识, 但仍需规范和引导。调查结果显示, 研究生在科研实践中能够

有较好的数据引用意识, 也能较好地认识到科研数据管理与共享阶段中涉及的数据归属权、知情同意权、合法商业利益、国家安全机密、隐私权等权益问题, 但在科研实践中如何遵循数据伦理与权益, 正确进行数据引用标识, 还需要进一步规范和引导。

[参考文献]

[1] Digital Curation Center [EB/OL]. [2016-06-24]. <http://www.dcc.ac.uk/training>.

[2] Education [EB/OL]. [2016-06-24]. <https://www.dataone.org/education>.

[3] 胡卉, 吴鸣. 国外图书馆数据素养教育最佳实践研究与启示 [J]. 现代情报, 2016 (8): 66-74, 78.

[4] NEWS& EVENTS [EB/OL]. [2016-06-24]. <http://data-archive.ac.uk/news-events>.

[5] Education [EB/OL]. [2016-06-24]. <https://www.dataone.org/education>.

[6] 顾立平. 科学数据开放获取的政策研究 [M]. 北京: 科学技术文献出版社, 2016: 5.

[7] Research data management guide [EB/OL]. [2016-07-20]. <http://www.cdrls.lib.vt.edu/dataguide/index.html>.

[8] Dissemination and Sharing of Research Results [EB/OL]. [2016-04-30]. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.

[9] 张晓林. 科学研究方法 [M]. 成都: 四川大学出版社, 1995.

[10] 顾立平. 科学数据权益分析的基本框架 [J]. 图书情报知识, 2014 (1): 34-51.

[作者简介] 胡卉 (1991-), 女, 硕士, 中国科学院文献情报中心助理馆员, 研究方向: 数据素养教育; 吴鸣 (1964-), 女, 硕士, 中国科学院文献情报中心研究馆员, 中国科学院大学经济与管理学院教授, 研究方向: 数据素养教育; 欧阳峥峥 (1983-) 女, 硕士, 中国科学院成都文献情报中心副研究馆员, 研究方向: 数据素养教育。

[收稿日期] 2017-10-09 [责任编辑] 菊秋芳