

面向社会科学研究的数字素养教育

——康奈尔大学经验与启示

屈宇

(南京大学图书馆, 南京 210023)

摘要: 科学研究范式的转变已对社会科学研究者的数据素养提出全新要求, 开展有针对性的数据素养教育存在紧迫性。在明晰相关概念与关系的基础上运用网络调研及内容分析方法, 全面梳理康奈尔大学面向社会科学研究的数字素养教育, 考察其教育架构、基础设施与资源建设、教育内容、服务策略等并总结其经验。针对我国现状, 为高校图书馆开展面向社会科学研究的数字素养教育提出建议: 重视研究范式转变, 提高数据素养培养意识; 推进多主体协作, 构建合理有效的数字素养教育体系; 加快建设数据基础设施和实践场所; 大力整合资源等。以期尽快提升社会科学研究者的数据素养, 适应研究范式转变, 助力社会科学的发展。

关键词: 数据素养; 数据素养教育; 社会科学研究; 康奈尔大学; 高校图书馆

中图分类号: G254.97 DOI: 10.3772/j.issn.1673-2286.2023.12.003

引文格式: 屈宇. 面向社会科学研究的数字素养教育: 康奈尔大学经验与启示[J]. 数字图书馆论坛, 2023(12): 21-31.

互联网技术的兴起带来了大数据技术的蓬勃发展, 促使科学研究的范式转向以数据为核心的数据密集型研究, 研究者管理和使用数据的能力——数据素养因而在研究中变得举足轻重, 数据素养教育亦成为新时代高等教育的一大命题。

数据素养的缘起与教育实践始于21世纪初的西方。我国学者也较早对欧美实践进行引介, 在数据素养教育的理论层面展开探讨, 不少代表性研究成果涌现。黄如花^[1]指出应根据数据生命周期流程图, 高质量发展数据发现、收集、分析、共享和保存能力。孟祥保等^[2]建议我国高校注意分层次开展数据素养教育活动并注重教学方式的整合。吴爱芝等^[3]、胡卉^[4]分别介绍北京大学图书馆以及中国科学院大学的数据素养教育实践, 为高校图书馆开展数据素养教育提供了参考或建议。总体而言, 数据素养教育的研究多基于高校图书馆这一独立主体; 教育内容多为信息素养教育的延伸, 数据内容零散、不成体系, 更缺乏区分度; 教育对象以高校学

生或教师群体为主, 鲜少关注科研人员。特别地, 学界对社会科学研究群体在数据密集型研究中面临的冲击缺乏关注, 在回应社会科学领域特定数据素养教育需求方面存在研究空白。在实践层面, 图书馆界已经积极开展数据素养教育, 但我国高校数据素养教育整体仍处于发展的初级阶段。广泛浏览我国985高校图书馆网站, 发现高校面向社会科学领域的数字素养教育内容稀缺, 仅北京大学图书馆^[5]、上海交通大学图书馆^[6]以及武汉大学图书馆^[7]等发布过少量社会科学数据技能或工具培训课程。产生这种现象固然有教育主体数据素养教育理念落后的主观因素, 但客观上, 数据素养教育涉及面广、软硬件要求高、教学内容更新快、实践性强, 高校图书馆作为唯一教育主体, 仅凭一己之力满足不同需求也实属有心无力。

目前无论是在学界研究还是业界实践层面, 我国高校的数字素养教育均未能深入社会科学研究领域。然而, 大数据时代科学研究范式的转变对社会科学研究

究环境带来的冲击可谓猛烈,新型数据密集型研究以多源异构数据为支撑,数据价值的实现高度依赖社会科学研究者数据意识、数据技能以及数据管理能力,而传统的社会科学研究者在这些方面普遍居于弱势,无力应对新型研究范式的挑战。因而,面向社会科学研究群体开展有针对性的数据素养教育显得尤为迫切。在这方面,欧美发达国家具有先行优势,一些知名的海外高校在数据素养教育的多主体协作、领域细分化、内容特色化等方面积累了丰富的经验与成功的策略,值得我国高校研究与学习,康奈尔大学(Cornell University)即为此中翘楚^[8]。为了帮助社会科学研究群体应对研究范式的数据化转变,康奈尔大学面向社会科学研究者开发了具有针对性的专门培训,辅以完备的基础设施以及创新的激励策略,为校园优秀的社会科学研究者提供了先进、多层次、可扩展的数据素养教育,对保障康奈尔大学全球领先的社会科学研究地位贡献不菲。鉴于此,本文以康奈尔大学为调研对象,采用网络调研的方法,收集其在开展面向社会科学研究的数据素养教育方面的相关信息,具体包括:康奈尔大学数据素养教育的整体架构与机构组成;由康奈尔数据服务组织(Cornell Data Services, CDS)^[9]总体规划与推进的覆盖数据全生命周期的管理与教育信息;康奈尔大学图书馆(Cornell University Library, CUL)^[10]系统为支撑科学研究提供的数据素养教育策略、培训与活动;康奈尔社会科学研究中心(Cornell Center for Social Sciences, CCSS)为社会科学研究者提供的专门数据培训与特别举措等^[11]。对调研所得内容进行分析,总结康奈尔大学围绕社会科学研究开展数据素养教育的经验,以期为我国高校开展同类教育提供参考与借鉴,从而帮助我国社会科学研究群体尽快提升数据素养,适应研究范式转变。

1 相关概念与关系

1.1 数据素养的内涵

随着数据环境和研究认知的变化,数据素养的内涵描述亦不断丰富。Schield^[12]将数据素养界定为根据科研需求从不同数据源中获取、操作和总结数据,并据此推断结论的能力,这一表述突出了数据素养与研究的密切关系。Prado等^[13]则将数据素养视为信息素养的组成部分,包含个人获取、理解、批判性地评估和管理

数据的能力,以及在使用数据的过程中遵守道德规范和诚信原则,且不仅限于科研领域,这一表述符合人们的普遍认知,即数据素养脱胎于信息素养。Mandinach等^[14]认为数据素养是理解数据和有效利用数据进行决策的能力,包括辨别、收集、组织、分析、总结以及优化数据的能力。郝媛玲等^[15]认为,数据素养高的个体具备数据意识和数据敏感性,能够有效且恰当地获取、分析、处理、利用和展现数据,并对数据具有批判性思维。简言之,数据素养实质上就是个体在数据收集、组织、处理、分析和利用过程中应具备的系统化综合能力。

数据素养也被认为是当前时代公民应普遍拥有的基本素质,其具体能力要求可参考欧盟于2016年6月发布的2.0版《数字素养框架》(The Digital Competence Framework for Citizens, DigComp),其将公民基本数据素养内容明确为:①具有一定的数据知识,能正确地浏览、检索和过滤数据;②具有评估数据的意识和能力,认识到某些因素可能导致数据偏差;③具有一定的管理数据的能力,如制作在线或动态表格、可视化展示数据、保证数据的可靠性、选择合适的数据存储方式、进行数据挖掘和数据组织、了解数据隐私等^[16]。

在科学研究领域,数据素养已被视为研究者不可或缺的核心素养。一般认为,从事知识生产与科学创新的科学研究者只有具有数据意识、数据知识、数据技能和数据思维等,才能科学地评估、推演和解释数据,实施决策,解决基于数据的研究问题^[17]。在数据密集型研究范式中,研究者需要从海量而丰富的数据中发现、提出问题、解决问题,最终释放数据的价值,故科学群体所需的数据素养内容可具体表述为:①具有数据发现、数据获取、数据分析与数据挖掘能力;②具有数据处理、数据存储、数据发布(或出版)、数据共享、数据引用等数据管理能力;③了解并遵守数据管理和应用中的法律法规以及数据伦理道德等。

1.2 大数据时代的社会科学研究与数据素养

社会科学旨在从不同的角度对人类社会的经济、政治、法律等进行分门别类或整体的考察研究,对人类社会的组织结构、功能作用、稳定机制、变迁动因等进行调查研究,以获得关于人类社会发展与运行的系统知识和理论,使人类得以更有效地管理社会生活。定量研

研究与定性研究是社会科学研究者进行事实性研究的两种基本范式。随着技术的飞速发展,数据渗入人类生活的各个方面,并开始被社会科学领域广泛应用。起初,研究者重视获取数据与存储数据,强调数据的完整性。后来,研究者为了最大化数据价值,转向对数据进行分析处理,社会科学研究进入数据密集型研究范式^[18]。无论是对于定性社会科学家还是定量社会科学家,新范式带来的变革都近于颠覆性。米加宁等^[19]认为,数据密集型研究范式改变了传统的假设驱动的研究方法,转向基于科学的数据挖掘研究方法,在预先占有大量数据的基础上,通过计算得出之前未知的理论。社会科学家陈云松等^[20]指出,大数据向外推动社会科学与自然科学(尤其是计算机科学)的融合,向内推动社会科学学科间的交流与对话,跨学科研究与交叉学科研究不断涌现,海量的数据和全新的数据特征使得定量研究与定性研究的资料获得与分析方法逐步趋同。数据密集型研究范式使社会科学领域的研究方法及其应用方式发生了根本性改变,从而对研究者的素养提出全新的要求。

综合国内外研究与实践,发现在数据密集型研究中社会科学研究者应具备的数据素养体现在以下方面:①具有数据相关知识,如了解数据的本质/形式/价值、常用的数据工具和存储平台、数据隐私、数据伦理、学术诚信与道德等。②具有获取数据的能力,如数据库检索、资源检索等。③具有数据管理技能,如数据生命周期管理、数据管理计划制定以及数据采集、处理、审核、清洗、转换、集成、脱敏等。④具有数据分析能力,包括应用软件和工具对数据进行分析,优化、改造或创新现有社会科学方法,如社会网络、主体建模、社会计算等。⑤具有展示、交流与应用数据的能力,如助力科研论文写作与投稿、进行专业热点追踪等^[21]。其中,数据管理、数据分析等被视为数据密集型社会科学研究者的必备技能。然而,有别于建立在实验或定量计算基础之上的自然科学,传统社会科学习惯使用观察法或调查法,社会科学研究者普遍缺乏深层次的数据挖掘能力、数据管理能力、数据共享和交流能力。因而,面向社会科学的数据素养教育需结合研究者的职业生涯、科研项目过程、数据生命周期来设计培养路径和内容,重点发展由数据分析、数据管理与算法支撑的数据素养,促使研究者改进科研行为方式与方法,以应对全新的研究环境和社会科学发展存在的问题。

2 面向社会科学研究的康奈尔大学数据素养教育

康奈尔大学创建于1865年,是享誉世界的研究型大学,在社会科学领域亦声誉卓著,劳工关系、教育、传媒、应用经济、酒店管理等学科一直是该校的传统强项。在文学理论和质性研究方面,康奈尔的学术成就全球顶尖。通过网络调研了康奈尔大学及其多个部门的网站,以CDS、CUL与CCSS为主,主要收集了2022年9月—2023年9月有关数据素养教育的信息并进行整合与呈现,从中可知:面对社会科学研究数据化、计算化趋势以及社会科学跨学科交叉研究的兴起,康奈尔大学一方面积极支持广泛的社会科学研究,鼓励多学科或跨学科部门参与社会科学研究;另一方面,积极整合校内机构与资源,搭建科学高效的教育架构,创新教学内容,并采用多种激励措施,为社会科学研究者开展了具有领域特色和显著成效的数据素养教育,帮助社会科学研究者全面提升数据素养。

2.1 多部门分工与协作的数据素养教育架构

康奈尔大学的数据素养教育由研究与创新副校长领导下的研究部统管,研究部下设CDS,全面规划数据服务,帮助师生在采用最佳实践和不同解决方案管理、存储和共享数据的过程中提升数据意识和掌握数据管理技能。CDS是一个协作性校园组织,由CUL、CCSS、康奈尔统计咨询部(Cornell University Statistical Consulting Unit, CSCU)、IT@康奈尔(IT@Cornell)、康奈尔高级计算中心(Center for Advanced Computing, CAC)以及生物技术资源中心(Biotechnology Resource Center, BRC)等共同组成。CDS于2011年前后开始工作,10余年来,整合推进全校的数据基础设施建设,帮助师生创建和实施数据管理计划,在研究活动的任何阶段提供数据管理服务,并联合各协作单位开展特色数据技能培训以提高师生的数据素养。

图1所示为康奈尔大学数据素养教育的基本架构。数据素养教育围绕研究数据展开,贯穿数据生命周期与科研生命周期。CDS的协作单位一方面以各自具有的功能服务教学科研;另一方面,也发挥各自知识优势提供技术培训,共同组成校园的数据素养教育内容。在CDS统一部署下多机构分工与协作是康奈尔大学数据素养

教育体系的一大特点,在这一体系中,具有共同目标的不同机构经过协调与组合,在某一领域创建更深入和细化的特色教育内容。在CDS的整体规划下,由CUL

保障资源,CCSS深入开发技能,协同构筑兼具广度与深度的、完整的、面向社会科学研究的数字素养教育体系。

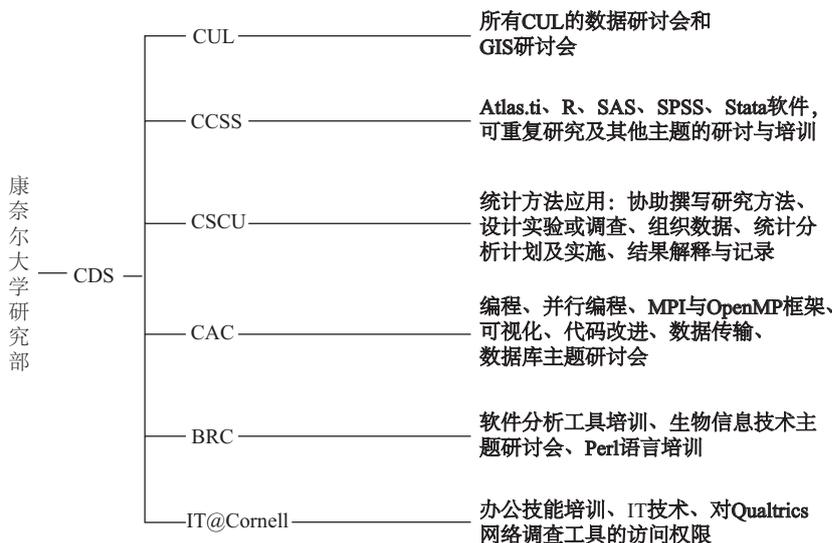


图1 康奈尔大学数据素养教育架构

2.2 CDS: 数据全生命周期的管理与教育

一个具有数据素养的研究者应该具备娴熟的数据运用与操作能力,如数据获取、挖掘、引用、共享、发布等,而对这些技能的学习与应用高度依赖数据的基础设施。CDS首先整合与发展了全校的数据基础设施,包括安全文件传输设施、9个以上先进的校内仓储以及数十个校外存储系统等,为社会科学研究者提供多个数据存储选项:①在康奈尔大学数字资源库eCommons中建立个人门户存储和分享数据;②在康奈尔大学箱式云存储Box中在线安全存储和共享研究成果,创建工作流程;③在定性数据库(Qualitative Data Repository, QDR)中管理和保存定性数据和多方法数据;④在其他若干特定主题的存储库中分享数据等。充足和发达的数据基础设施为社会科学研究者学习和实践数据管理技能提供了必备条件。

同时,CDS广揽来自科学、政策、数据、信息技术等领域的专家组成咨询委员会,以便满足所有康奈尔研究团体,包括社会科学研究团体的数据管理需求。CDS提供覆盖数据全生命周期的咨询服务,内容包括协作工具使用、数据收集和分析、数据监管、数据共享、数字人文、高性能计算、知识产权和版权、元数据描述、数据安全与隐私、数据存储与备份等。无论研究人员在研

究的哪个阶段遇到困难,都可以从CDS获得及时和专业的帮助,CDS通过个性化的咨询服务提升研究人员的数据管理能力。

此外,CDS还常年组织数据管理的自定义培训,如表1所示,内容包含数据管理每一个阶段必备的知识和技能。绝大多数培训为初级或中级难度,有助于减轻学习者的畏难情绪,激发学习热情^[22]。总之,CDS围绕研究数据管理开展的服务为整个康奈尔大学的数据素养教育奠定了优越的硬件基础并普及了数据管理知识和技能,而后者是研究人员数据素养中不可或缺的部分。

2.3 CUL: 服务大学研究的数据素养教育

图书馆素来是大学数据素养教育最重要的主体。CUL由位于纽约伊萨卡的主馆和21个特色分馆共同组成,是美国十大学术研究型图书馆之一。作为康奈尔校园数据素养教育的重要场所之一,CUL的数据素养教育以研究为中心,这一点有别于加拿大达尔豪斯大学以学生为主导^[23]、美国南佛罗里达大学以教师为中心^[24]的数据素养教育。康奈尔大学认为校园里每一位教师都同时是一位研究者,无论他手头有没有研究项目;鼓励每一位学生参与研究,哪怕是本科生,也被视作潜在的

表1 CDS定制培训菜单

阶段	课程内容	时长/分钟	级别
前菜	学习使用开放科学框架 (Open Science Framework, OSF)	30	初级
	保存实验笔记本的最佳实践	30	初级
	数据可视化介绍	30	初级
	文本分析概论	30	初级
主菜	将QGIS软件应用于科学、人文和社会科学	60	中级
	使用OpenRefine工具清理杂乱的数据	60	中级
	如何使用Twitter数据	60	初级
	使用Excel进行数据分析: 公式与函数	60	初级
甜品	如何使你的数据可再现	45	初级
	使用Tableau软件可视化数据和开发数据	45	中级
	数据技能: 如何在电子表格中组织数据	45	初级
	数据管理计划新手指南	45	初级
副菜	使用Excel透视表	60	初级
	使用Excel宏	60	中级
	什么是元数据以及为何要使用它	30	初级
	如何分享与发布数据	60	初级
	数据技能: 维基百科编辑	45	初级
	如何使用NVivo软件来做定性数据分析	60	高级

研究者。在这一理念指导下, CUL以科学研究者的数据素养为标准, 经由资源运用、空间实践以及培训活动等推进数据素养教育。

2.3.1 资源整合与空间建设

资源是研究的基础, 发现和使用资源是研究者必备的能力, 故CUL将馆藏资源及全校课程资源进行整合, 并在网站首页创建研究指南作为数据素养教育一站式资源导航, 供师生按课程、主题或研究类别检索资源, 学习数据知识, 培养研究技能^[25]。研究指南分为三部分。①课程指南。包含87类学科课程, 其中一半以上为人文社科类, 每类下另有细分, 课程种类极其丰富。②主题指南。共61类主题导航, 约2/3为人文社科主题, 每大类再作细分。③研究入门指南。即资源或技能指南, 专为集中某一特定研究领域下的各种资源或必备技能而设, 品类丰富, 可提供类目和人名检索服务, 如: 点开“新闻研究指南”, 用户可快速链接到新闻文章, 或者免费在线获取新闻资源等; 点开“数据可视化指南”, 则所有数据可视化方法、工具、资源可一网打尽。研究指南为研究人员开发研究课题、查找研究资料、评估信息来源和进行引文管理等奠定了资源基础, 有助于研究人员在使用过程中发展数据检索能力、评估能力、研究技能和数据意识等重要的数据素养。

CUL还组建馆员服务团队, 应社会科学研究之需建设特定数据源。以商业与创业研究为例: 馆员团队以专业知识和资源帮助研究者探索、理解和查找研究相关的各种主题的数据, 提供面谈、邮件或在线形式的服务, 定制培训课程与研讨会, 创建资源导航, 汇集各类文章、研究报告和统计数据等, 采集所有相关学术研究数据, 并提供彭博、路孚特及RS Means数据库的校园访问权限, 为商业与创业研究提供海量且不断迅速增长的数据资源。建立特色数据源是CUL嵌入式数据素养教育的重要举措, 体现了数据素养教育的精细化水平。

CUL同样很重视为社会科学研究提供空间支持, 以便社会科学研究者磨炼对数据技术的实操技能。以建于19世纪60年代的欧林图书馆为例: 这座古老的人文社科类分馆建设了最现代化的数字实验室CoLab^[26], 支持人文社会科学的前沿实验性研究。CoLab提供与数据课程配套的实训设备资源与演练环境, 以一对一咨询、研讨会和课堂教学的形式帮助师生掌握数据监管、数据出版、数据可视化或计算文本分析等技能。特别地, 为了鼓励人文社科领域的博士生加入学习, CUL还实施了奖学金计划, 赞助相关学科的博士生来到实验室与馆员、同行或专家一起, 在CoLab以沉浸式阅读和讨论的方式共同探索数字化的研究与教学方法, 获得开发、实施和管理复杂人文社科项目的实践经验。数字

实验室为社会科学研究者提供了数据素养教育实践空间,有助于他们掌握新兴数据技术以及未来赢得更广泛的学术影响。

2.3.2 数据素养培训课程与活动

CUL围绕科学研究设计了丰富的数据素养教育课程与活动,无论是写作入门培训还是专业领域的科研项目,馆员都能就各种研究主题提供资源建议与搜索策

略;专业馆员还嵌入学术部门或研究项目,通过课程、中小型研讨、演讲、竞赛等形式开展数据素养教育。表2列举了CUL开展的主要培训与活动。

由表2可见,CUL的培训课程与活动内容非常广泛,不仅兼顾了自然科学与社会科学研究所需的基本的数据管理与处理技能,还普及应知晓并遵守的法律法规、科研伦理等重要内容,并引导人们认识开放科学。通过与研究相关的全面教育,促使研究者的数据意识、数据技能和数据思维获得全面提升。

表2 CUL数据素养教育培训与活动(部分)

课程名称	课程内容	形式	教育目标
数据隐私素养培训	了解互联网运作;识别潜在的隐私风险;掌握减少隐私风险的实际行动	公开讲座、定制研讨会	了解、评估和保护数据隐私
研究安全与道德	学习特定研究领域的标准做法、安全措施、程序与法律等	强制培训	普及学术诚信与数据伦理
科学研究入门	制定研究问题;确定关键词检索数据;评估、区分和分析数据源;识别错误虚假信息	咨询与讨论	了解研究数据的检索与评估方法
写作2100	人文社科研究工具和方法	课程	掌握研究基本技能
Scrivener入门	写作软件Scrivener的功能及使用	研讨会	掌握研究基本技能
引文管理:来源、工具及软件	引用来源指南;引用工具ZoteroBib、引文管理软件Zotero/Mendeley/EndNote使用方法	研讨会、咨询	普及数据检索与书目组织方法
数据的本质	了解数据平台、设备、标准和机构的作用	面谈、活动	了解数据环境
创建数据管理计划	学习编写成功的数据管理和共享计划	课程、研讨会	学习数据管理计划
桌面地理信息系统QGIS介绍	基本任务:加载数据、更改数据样式、安装插件、分析以及导出地图图像	课程、研讨会	学习数据分析技能
研究者工具包	搜索、整理和组织数据,管理数据文件;存储数据、处理电子表格数据以及优化数据包;数字隐私等	研讨会	掌握研究基本技能、提供职业生涯指导等
爱数据周	数据访问、分析、发现、管理、共享和保存;从编码到数据管理;找到数据培训资源等	研讨会、系列交流活动	学习数据管理、数据交流与应用技能
开放学术:开放获取、开放科学、开放数据	开放科学运动;开放性倡议(预印本、注册报告、永久标识符和社区参与平台)	小组讨论、课程、研讨会	宣传开放科学,促进研究过程的开放
开放获取周	资助研究的途径;arXiv的变化及影响力;开放获取出版基金;康奈尔可持续发展等	系列演讲活动	推广开放获取出版、预印本平台等

CUL从研究角度切入数据素养教育,并以科学研究所需的数据素养为教育标准,彰显了研究型图书馆在数据素养教育方面的高远立意;同时,CUL为开展数据素养教育整合资源、升级空间、开发课程与组织活动,亦极大地体现了图书馆的价值。

2.4 CCSS:全面先进的社会科学研究支持

如果说CUL侧重为科学研究者提供通识性数据素养教育,那么CCSS则专注于在社会科学研究领域通过硬件支持、措施激励、专门培训等进一步打磨科学研究者数据技能。CCSS成立于2004年,起初主要通过提供先进的计算基础设施、基金和奖学金来支持社会

科学的定量与定性研究。近年来,随着社会科学研究范式的数据化,CCSS加大了对社会科学研究者的培训力度,成为数据素养教育的主力。

2.4.1 提供面向社会科学研究的专门基础设施

基础设施是开展数据素养教育的必备条件,CCSS在康奈尔大学强大的基础设施之上又为社会科学研究特别提供高性能云计算设施,为定量和定性软件提供安全、可扩展和可定制的环境。为了支持社会科学知识的积累,确保研究与数据的可获得性,CCSS提供了多个数据存储选项,并把关数据的完整性、安全性与合规性,帮助研究者找到管理数据的最佳方式。对于不能

在云计算环境中使用的机密、受限、敏感或专有数据，CCSS帮助研究者使用安全数据计算环境并实现远程访问。CCSS提供的专用数据基础设施以及配套服务为研究者提升数据素养提供了机会和保障，并进一步强化了社会科学研究者的数据管理技能，使他们更加适应云计算和安全存储的要求。

CCSS高度重视社会科学研究者数据技能的实践，认为随着大数据的出现，工具、基础设施和思考方式也应随之扩展，研究人员需要在设计和使用数据科学方法方面不断实践以发现数据潜藏的价值。因此，CCSS着力汇聚校园内近40个不同研究主题的社会科学数据实验室，鼓励社会科学研究者走进实验室，通过沉浸式学习或动手操练来实际探索数据，发现数据如何与社会相互影响，从而提升研究者从事社会科学和数据科学交叉研究的能力。

2.4.2 促进社会科学和数据科学加速融合

数据密集型研究范式极大地改变了社会科学领域的研究方法，数据收集与数据分析技术变得尤为关键。鉴于传统的社会科学研究者客观上普遍缺乏数据分析技能，主观上又存在明显的畏难情绪，在数据素养教育中实施有效的策略和激励手段就显得很有必要。CCSS与其下辖的众多研究机构采用多种支持手段，如提供经费资助、创造跨学科合作与交流机会、增加培训研讨等，激发社会科学研究者学习数据技能的热情，推进社会科学传统研究方法与数据科学的融合或转化。

以定性研究为例，为了加强对定性科学家的支持，CCSS设立了定性与解释性研究所QuIRI，汇集校园内讲授、使用和开发定性研究方法的研究人员，以多种形式加强社会科学定性研究方法的培训。①每月开展定性方法系列研讨会，探讨与定性研究相关的方法、技术和研究项目，涉及很多数据技能研讨。②开设教师培训学院，在春季、暑期、秋季讲授定性研究方法。③设立教师工作组，为使用定性研究方法的社会科学研究者就特定的方法或软件提供嵌入式培训^[27]。

鉴于大数据时代算法技术不断升级，传统量化研究逐渐走向计算社会科学，这对研究人员提出了更高的数据技能要求。为此，CCSS企业与社会数据科学中心推出三大项目来加强社会科学研究中快速发展且日益重要的数据科学研究，促进算法技术与计算社会科学的融合。①研究助理教授项目。每年遴选3名对数据

科学感兴趣的年轻科学家，通过3年的探索、交流、合作与碰撞，深度发展数据技能，解决社会科学前沿问题。②新兵训练营。每学期开学前举行为期1周的密集会议，汇集统计和数据科学、运筹学和信息工程、信息科学、计算机科学、电子和计算机工程、经济学以及数学系的教师，围绕一系列社会科学领域的研究计划进行探讨合作，创新数据科学研究，以解决该领域的关键问题。③为数据科学博士生在社会科学研究方向开设辅修专业^[28]。

CCSS大力促进社会科学和数据科学的加速融合，为校园社会科学群体营造了积极学习数据科学的氛围，激发了他们深度学习数据技能的热情，促进了社会科学的发展与变革。此外，CCSS也很注重挖掘更多的数据人才来从事数据素养培训，并设立了数据科学奖学金^[29]。数据科学奖学金的目标是为受到资助的康奈尔博士生提供数据科学研究员职位，这些数据科学研究员可以来自任何科系，需要胜任下述任何一项工作。①开发和创建数据技能研讨会。研究员每学期发起几次研讨会，讲授数据科学技术并提供相关咨询服务。②开发和运营云计算基础设施。主要通过对社会科学家的软件使用趋势进行分析和建模，帮助优化资源分配来支持以云计算为重点的计算社会科学研究。③组织和策划活动，宣传数据科学。研究员可通过策划项目、主办会议或调查采访来帮助扩展康奈尔大学的数据科学社区，让更多的社会科学家参与计算工作。数据科学奖学金项目可谓一举多得：首先，直接促进了数据科学研究员自身的学业和职业发展；其次，由于数据科学研究员有能力胜任数据技能培训，显著缓解了CCSS的师资压力；最后，奖学金项目还具有很好的宣传作用，扩大了数据素养教育的受众。

2.4.3 同步新兴技术发展的数据培训课程

数据素养培训是数据素养教育的重要组成部分，毕达天等^[30]基于SEM及fsQCA方法对科研人员数据素养的影响因素进行综合分析，结果表明教育培训已成为科研人员数据素养最重要的影响因素。CCSS高度重视对社会科学研究者的数据培训，每月举办的数十场社会科学主题相关的讲座、研讨会和会议中，相当一部分为数据素养培训，培训团队由培训师、高级数据科学研究员和数据科学研究员共同组成，培训形式包括线上、线下研讨会。CCSS研讨会规模一般不大，讨

论的话题相对集中, 互动性很高, 学习效果出色。2020年以前, CCSS的培训以数据知识和数据管理为主, 近几年随着数据管理的逐渐普及以及社会科学研究的加快转型, 培训重点转向数据分析技能、数据计算工具

与软件。CCSS推出了大量针对社会科学研究的深度培训, 内容紧跟融合社会科学与数据科学的最新技术。对CCSS的培训进行回溯, 剔除部分重复课程, 分类汇总CCSS数据素养教育课程, 见表3。

表3 CCSS数据素养教育课程(部分)

课程主题	课程名称
研究与计算	关于CCSS数据访问服务
	在康奈尔大学使用受限数据
	在社会科学研究中使用应用程序接口
	访问康奈尔联邦统计研究数据中心获取可用数据
	在线查找开放数据集: 纽约开放数据、Census.gov等
	RMarkdown入门
	用RStudio/RMarkdown文档创建报告
	CCSS计算资源介绍
定性数据分析	社会科学家的计算工具
	使用Atlas.ti进行定性分析
	使用MaxQDA进行定性分析
	使用NVivo作为研究工具
统计数据分析	定性分析实践会议: Atlas.ti、MaxQDA
	SPSS系列: SPSS入门、SPSS速成、中级SPSS
	R系列: R入门、R包生态系统、R中的数据可视化
	Python系列: Python入门、Python包、Pandas、Python高级语言与接口
	Stata系列: Stata入门、Stata中级操作、Stata高级编程
	SAS入门
	结构方程建模
	自然语言处理入门
	使用Qualtrics联合分析个人潜在偏好
	OpenRefine: 数据清洗、转换及扩展
数据操作与分析实践(为具有入门知识并希望进一步磨炼编码技能的研究人员设计)	Tidyverse数据清洗入门
	网络抓取/应用程序接口系列: 提取数据、变量分析、过滤数据
	Python系列实践课程
	使用Python库的Beautiful Soup和Selenium实现Web抓取
	R系列实践课程
	Stata系列实践课程
	机器学习入门
	机器学习: Python中的监督学习
高级编程	机器学习: Python中的无监督学习
	机器学习: 自然语言处理(Python)
	机器学习: R中的监督学习
	R中的因子分析
	在社会科学研究中应用机器学习模型
	Python中的Web抓取(Beautiful Soup)
	Python中的Web抓取(Selenium)
	地理空间分析: R中的映射
	地理空间分析: 在R中绘制普查数据
	地理空间分析: 创建交互式地图和数据可视化
研究数据管理: 发布、存档与重现	如何让研究更透明且可重复
	为发布、存档和共享准备高质量的数据和代码
	Qualtrics在线调查: 调查流程、逻辑、嵌入数据、触发器、配额
	R中的发布就绪表
	Stata中的发布就绪表
	定性数据仓储
	使用GitHub进行版本控制

由表3可见,数据分析是CCSS研讨会的主要议题,同时,数据管理技能与研究素养也获得重视。与CDS以及CUL开展的数据素养培训相比,CCSS的培训更像是站在巨人的肩膀之上:在CUL夯实广泛的研究能力和基本数据素养、CDS强化数据管理技能的基础上,CCSS将社会科学研究者引向更专业、更精深与更前沿的方向。三者协同,为康奈尔大学社会科学研究者提供全面、新颖且富有针对性的数据素养教育内容,助力社会科学研究者为世界更多更紧迫的社会与行为问题提供解决方案。

2.5 康奈尔大学数据素养教育经验总结

通过对CDS、CUL以及CCSS数据素养教育策略与内容的梳理,可以总结出康奈尔大学面向社会科学实施的数据素养教育具有以下特点。①理念先进。专为社会科学研究者开展数据素养教育,表明领导层对人类社会发展趋势与社会科学研究环境的深刻理解。当研究范式发生转变时,高校和研究者都应该积极应变,而非消极等待。②管理灵活。这充分体现在数据素养教育体系的架构上:设立CDS进行总体规划,围绕提升师生数据素养这一目标,部署组织内各机构分工和协作,从而形成一个科学、完备、集约的数据素养教育体系。由此可知,开展数据素养教育尤其是领域特色化教育,仅靠图书馆单打独斗是不现实的,由图书馆联合校内相关机构才是明智的选择,而这依赖管理层面的规划与协调。③强大的基础设施与资源整合为数据素养教育奠定了坚实的基础。在海量的数据资源中发现和获取数据,在数据实验室中对数据进行分析和处理,以及通过数据基础设施传输、存储与发布数据,都离不开数据资源与基础设施。如果说数据收集、分析与管理技能是数据密集型社会科学发展的基石,那么基础设施与数据资源则是研究者发展数据素养能力的重要基石。④多样化的教育服务与策略激发了社会科学研究者的学习热情。很多社会科学研究者怯于学习和运用数据技能,主动的服务和激励措施有助于提高研究者的学习积极性。无论是广泛的咨询服务、嵌入式的个性服务,还是各种名目的奖学金,都大大提升了社会科学研究者掌握新兴数据技能的信心,推动他们尽快适应新变化。⑤全面和新颖的培训内容促进了社会科学研究者数据素养的快速提升。康奈尔大学为社会科学研究的培训内容主要由CDS的数据管理内容,

CUL的涉及研究技能、数据管理与数据分析的通识性内容,以及CCSS以数据分析为主的专门性内容共同组成,全面、具有层次性和梯度、易于模块化组合,且更新及时,始终关注新技术,有利于社会科学研究者掌握最新的技术,开展最前沿的社会科学研究。

3 康奈尔大学数据素养教育实践对我国的启示

大数据时代,数据已成为社会发展的主要驱动力,开展数据素养教育乃应时之需。美国大学与研究图书馆协会(Association of College and Research Libraries, ACRL)指出,研究型图书馆集资源、能力、服务经验于一体,是提供数据管理服务和开展数据素养教育的最佳主体,能够帮助研究人员应对数据驱动研究环境中的挑战^[31]。然而,从国内外数据素养教育实践来看,以单一图书馆为主体的教育实践效果不佳。我国高校图书馆在发展意识、基础设施、实践手段、教育人才、教育内容等方面都存在严重制约,较难开展面向社会科学领域的的数据素养教育。但是,社会科学研究范式转变带来的冲击已使对数据素养教育的需求变得迫切。他山之石,可以攻玉,借鉴康奈尔大学为社会科学研究实施的数据素养教育,我国高校图书馆可在以下方面作出努力。

(1) 重视研究范式转变,提高数据素养培养意识,积极规划面向社会科学的数据素养教育。社会科学的环境已不同以往,数据科学与社会科学的渗透融合使得对社会科学研究者的数据素养要求向专、精、深、新发展,教育主体应充分认识这一变化。一方面,加强数据服务与激励,采用创新策略如研究资助等激励社会科学研究者积极学习数据技能,或完善数据服务以使研究者在良好的体验中提升数据素养;另一方面,要主动为社会科学研究寻求具有针对性和扩展性的数据素养教育内容,尤其应高度关注新技术发展,不断更新教育内容,帮助社会科学研究者以最佳的数据素养实现研究突破。

(2) 推进多主体协作,构建合理有效的数据素养教育体系。在CDS领导下的多机构协作模式集约、科学、富有成效,值得借鉴。我国高校也应通过制定整体发展规划,加强协调,推进图书馆与社会科学、数据科学以及计算机信息科学部门的合作,共同创建适合社会科学研究的软件、工具和培训内容,提高社会

科学工作者的数据技能。此外,高校图书馆亦可效仿CCSS的数据科学奖学金政策,吸纳具有专业技能的博士生加入培训师队伍,充实数据素养教育师资。

(3) 加快建设数据基础设施和实践场所。数据管理、数据处理和分析都是操作性很强的技能,研究者需要在具备相应硬件和数据的环境中学习操练方可熟能生巧。数据密集型社会科学研究高度依赖基础设施,研究者需要数据仓储来学习和开展数据全生命周期管理,还需要数据实验室来模拟研究数据环境。目前我国仅北京大学图书馆、复旦大学图书馆等少数图书馆能运行数据平台,实现数据的管理;另一些图书馆建设了机构知识库,但多数存而不享;更多的图书馆还不具备数据存储条件,使研究者无从实践数据管理技能。数据实验室在高校也不够普及。高校应积极争取建设资金、技术和人力,加快完善数据存储设施和数据实验室建设,为社会科学研究者提升数据素养提供实践平台。

(4) 图书馆应大力整合资源,为数据素养教育提供支撑。数据资源是高校图书馆开展数据素养教育服务的利器,也是图书馆最独特的优势。缺乏资源的数据素养教育有如无源之水。图书馆应加大社会科学数据资源建设力度,利用购买、收割或开放获取等方式充实自身储备;同时,收集整理全校相关资源,建立图书馆自己的研究指南,为数据素养教育提供一站式数据资源平台。

4 结语

康奈尔大学面向社会科学研究的数据素养教育科学、先进、高效,有利于快速全面提升社会科学工作者的数据素养,助力数据密集型社会科学研究。随着信息革命的深入,社会科学领域正涌现越来越多的跨学科研究和交叉学科研究,社会科学研究者也需要持续的、不断更新的数据素养教育。我国高校图书馆应提升理念,积极应变,注重整体规划协作、加强基础设施建设、创新服务手段、追踪新技术发展,推动数据素养教育向更深入、更专业和更完善的方向发展,促进社会科学研究转型与发展。

参考文献

[1] 黄如花. 面向高质量发展的数据素养教育[J]. 图书馆建设, 2020

(6): 26-29.

- [2] 孟祥保, 符玉霜. 美国数据素养课题剖析与启示[J]. 图书与情报, 2018 (5): 126-133.
- [3] 吴爱芝, 王盛. 高校图书馆数据素养教育体系设计研究: 以北京大学图书馆为例[J]. 大学图书馆学报, 2020, 38 (6): 96-103.
- [4] 胡卉. 图书馆数据素养教育服务模式设计: 以中国科学院大学为例[J]. 图书馆理论与实践, 2018 (9): 85-89.
- [5] 北京大学图书馆. 一小时讲座[EB/OL]. [2023-10-17]. <https://www.lib.pku.edu.cn/4whjy/41xszjy/index.htm>.
- [6] 上海交通大学图书馆. 专题培训[EB/OL]. [2023-10-17]. <https://www.lib.sjtu.edu.cn/f/content/detail.shtml?id=6128&lang=zh-cn>.
- [7] 武汉大学图书馆. 学分课程[EB/OL]. [2023-10-17]. https://www.lib.whu.edu.cn/webfile/category/Credit_courses/305.html.
- [8] Cornell University. Research[EB/OL]. [2023-10-14]. <https://www.cornell.edu/research>.
- [9] Cornell University. Cornell Data Services[EB/OL]. [2023-10-13]. <https://data.research.cornell.edu>.
- [10] Cornell University Library. Research support[EB/OL]. [2023-10-14]. <https://www.library.cornell.edu/research-teaching/research-support>.
- [11] Cornell University. Cornell Center for Social Sciences[EB/OL]. [2023-10-14]. <https://socialsciences.cornell.edu/about>.
- [12] SCHIELD M. Information literacy, statistical literacy and data literacy[J]. Iassist Quarterly, 2004, 28 (2): 7-14.
- [13] PRADO F J C, MARZAL M. Incorporating data literacy into information literacy programs: core competencies and contents[J]. Libri: International Journal of Libraries and Information Services, 2013, 63 (2): 123-134.
- [14] MANDINACH E B, GUMMER E S. A systemic view of implementing data literacy in educator preparation[J]. Educational Researcher, 2013, 42 (1): 30-37.
- [15] 郝媛玲, 沈婷婷, 高珊. 高校数据素养教育实践的思考和建设: 基于哈佛大学案例和我国图书情报人员访谈的分析[J]. 图书情报工作, 2015, 59 (12): 44-51.
- [16] 汪庆怡. 从欧盟数字素养框架(DigComp2.2)论全民数字素养的提升[J]. 图书馆杂志, 2023, 42 (3): 97-106.
- [17] 赵芳. 国内数据素养教育研究脉络与展望[J]. 图书馆理论与实践, 2023 (4): 30-40, 46.
- [18] 郑一卉. 范式与方法: 对大数据环境中社会科学研究的反思[J]. 新闻知识, 2016 (1): 12-14.
- [19] 米加宁, 章昌平, 李大宇, 等. 第四研究范式: 大数据驱动的社会

- 科学研究转型[J]. 社会科学文摘, 2018 (4): 20-22.
- [20] 陈云松, 吴青熹, 黄超. 大数据何以重构社会科学[J]. 新疆师范大学学报(哲学社会科学版), 2015, 36 (3): 54-61.
- [21] 章昌平, 米加宁, 李大宇. 数据科学研究在社会科学中的应用前景[J]. 社会科学, 2018 (9): 78-88.
- [22] Cornell Data Services. Workshops[EB/OL]. [2023-10-13]. <https://data.research.cornell.edu/workshops>.
- [23] 陈媛媛, 王苑颖. 加拿大数据素养教育实践及启示: 以达尔豪斯大学为例[J]. 情报理论与实践, 2019, 42 (6): 166-171.
- [24] 黄茂汉, 邱瑾. 美国南佛罗里达大学全学科数据素养教育体系构建及启示[J]. 图书馆工作与研究, 2023 (3): 46-55.
- [25] Cornell University Library. Cornell University Research Guides[EB/OL]. [2023-10-13]. <https://guides.library.cornell.edu/home>.
- [26] Cornell University Library. Olin & Uris Libraries[EB/OL]. [2023-10-14]. <https://olinuris.library.cornell.edu/public-computing-and-technology>.
- [27] CCSS. QuIRI[EB/OL]. [2023-10-14]. <https://socialsciences.cornell.edu/funding-and-programs/qualitative-interpretive-research-institute>.
- [28] CCSS. Center for Data Science for Enterprise and Society[EB/OL]. [2023-10-14]. <https://datasciencecenter.cornell.edu>.
- [29] CCSS. Data Science Fellows Program[EB/OL]. [2023-10-14]. <https://socialsciences.cornell.edu/funding-and-programs/data-science-fellows-program>.
- [30] 毕达天, 曹冉. 科研人员数据素养影响因素分析: 基于SEM及fsQCA方法[J]. 情报学报, 2021, 40 (1): 11-20.
- [31] ACRL. Annual Report 2020-2021[EB/OL]. [2023-10-17]. <https://crln.acrl.org/index.php/crlnews/article/view/25246/33117>.

作者简介

屈宇, 女, 硕士, 馆员, 研究方向: 科研数据管理、数据素养教育, E-mail: 987105895@qq.com。

Data Literacy Education for Social Science Research: Experience and Revelation from Cornell University

QU Yu

(Nanjing University Library, Nanjing 210023, P. R. China)

Abstract: The change of scientific research paradigm has put forward new requirements for data literacy of social science researchers, and it is urgent to carry out targeted data literacy education. On the basis of clarifying relevant concepts and relationships, this paper uses network research and content analysis methods to comprehensively review the data literacy education implemented by Cornell University for social science research, investigate its educational structure, infrastructure and resource construction, educational content, and service strategy, and sum up its experience. According to the current situation of our country, this paper puts forward some suggestions for university libraries to carry out data literacy education for social science research. They should pay attention to the change of research paradigm and improve the awareness of data literacy training; promote multi-subject collaboration to build a reasonable and effective data literacy education system; accelerate the development of data infrastructure and practice sites; and integrate resources vigorously. This paper aims to improve the data literacy of social science researchers as soon as possible, adapt to the change of research paradigm, and help the development of social science.

Keywords: Data Literacy; Data Literacy Education; Social Science Research; Cornell University; University Library

(责任编辑: 王玮)