



促进开放科学:国外高校图书馆 2022年“爱数据周”活动分析与启示

□杨鹤林*

摘要 开放科学日益受到关注,随着科研对数据的依赖不断增强,数据服务已成为高校图书馆促进开放科学的重要手段。本文通过案例研究、内容分析及文献综述,对2022年“爱数据周”活动中的136项高校图书馆活动的基本要素及内容特征进行了调研和分析,并从组织者、活动内容策划及活动形式等方面对国内相关工作提出建议。

关键词 高校图书馆 开放科学 开放数据 爱数据周 数据服务

分类号 G250

DOI 10.16603/j.issn1002-1027.2023.03.006

目前,科学本身的多样化、快速的技术进步及海量科研数据的产生促成了现代意义的“开放科学”大发展^[1]。尽管人们对开放科学有多种阐释,但均对开放数据在其中发挥的驱动作用高度认同。欧盟开放科学战略的首要目标便是开放数据^[2];经济合作与发展组织在《使开放科学成为现实》的报告中指出,数据对科学的驱动日益显著,各国要完善数据政策、技能、立法、共享等方面的建设^[3];联合国教科文组织于2021年11月通过了《开放科学建议书》,意味着在开放科学活动中,依据FAIR、CARE等原则对数据及元数据进行使用和管理已成为全球共识^[4]。长久以来,高校图书馆紧随科研范式和学术信息体系的发展变化不断创新服务,在推动开放科学、开放数据的运动中发挥了积极作用^[5]。

1 研究背景

笔者以“university librar*”“academic librar*”“open science”“open data”“data services”“RDM”等为检索词,在EBSCO-LISTA数据库以及互联网检索外文学术文献、相关国际组织和图书馆协会发布的文件和报告。经筛选后得到60余篇研究开放科学与高校图书馆数据服务的相关文献。总体来

看,国外高校图书馆已经通过多种渠道对开放科学理念和数据服务进行了推广,并建立了相关社团。推广的主要路径包括:(1)利用成熟的线下用户教育业务,如纽约大学图书馆的“数据服务实验室”^[6]、亚利桑那大学图书馆的“R语言开放实验室”^[7]等;(2)利用Libguide、社交媒体等线上信息渠道,如康涅狄格大学图书馆的科研数据管理 Libguide^[8];(3)利用竞赛型数据产品,如普渡大学图书馆的数据马拉松^[9],而其中影响力最大的是安德鲁·W·梅隆基金会资助的“档案释放”(Archives Unleashed)项目,该项目源于2016年多伦多大学图书馆举办的数据马拉松活动,2023年赛事是探索“档案释放”云与“互联网档案馆”网站(Internet Archive)的 Archive-It 项目内容整合解决方案,约克大学图书馆是主办方之一^[10]。

上述案例已能大致勾勒当前国外高校图书馆的开放数据推广路径。若能进一步考察案例集中的相关大型活动,则更利于我们快速了解图书馆界的最新实践及工作思路。目前国际上影响较大的数据主题大型活动包括:“国际开放数据日”(International Open Data Day)、“地理信息系统日”(GIS Day)及“爱数据周”(Love Data Week)等。其中“爱数据周”活动经过6年的发展,已成为一项由高校图书馆主

* 杨鹤林,邮箱:80943694@qq.com。



导的、稳定的大型数据主题活动，其目的是“提升数据意识、创建数据社区，以此推进研究数据的管理、共享、保存、再利用工作和立足于图书馆的数据服务不断深化”^[11]。2022 年“爱数据周”活动主题是“数据属于大众”^[12]，带有明显的开放科学色彩。通过考察该活动网站及各参与机构的相关信息，共获得 152 项活动案例，其中 136 项分别由 36 所高校图书馆主办，本研究即以此 136 项活动为调查对象，分析国外高校图书馆面向开放科学的数据服务基本要素及其整体特征，以期为我国高校图书馆提供参考借鉴。

2 基本情况及其特征分析

2.1 北美高校图书馆是最主要的活动组织者

从主办活动的高校图书馆地理分布及推广活动数量看，北美地区是“爱数据周”活动最密集的举办区域，具体情况见表 1。其中加州大学图书馆是最为积极的主办者，共在 9 个校区推出了 30 项活动。

表 1 高校图书馆的地理分布及推广活动数量

国家	高校图书馆		推广活动	
	数量(所)	占比(%)	数量(项)	占比(%)
美国	18	50	90	66.2
加拿大	13	36.1	34	25
英国	4	11.1	11	8.1
法国	1	2.8	1	0.7
合计	36	100	136	100

2.2 顶尖高校图书馆示范效应明显，普通高校图书馆积极响应

我国关于国外高校数据服务的相关研究主要聚焦于顶尖高校，如夏琬钧等研究了 QS 排名前 50 的高校^[13]；李正超聚焦康奈尔大学、斯坦福大学、剑桥大学、牛津大学等 10 所世界著名高校^[14]；王利君以英国 QS 排名前 10 的高校为样本^[15]等。笔者采用 2022 年 QS 世界大学排名（共 1300 所高校参加排名）对本研究相关的高校进行了分类，结果见表 2。可以发现，一方面顶尖高校图书馆确实发挥了明显的示范效应，9 所排名前 10% 的机构举办了近半数的活动（42.4%）；另一方面近四成（38.9%）主办机构是排名后 50% 以及无排名的高校，贡献了超过三分之一的活动（36.4%）。可见国外高校图书馆对开展数据服务普遍持积极态度，这有利于开放科学理念的进一步扩散，也使得数据服务能较普遍地下沉到

不同层级的高校中去。

表 2 按 2022 QS 世界大学排名的高校图书馆数量及活动场次

2022 QS 世界大学排名	高校图书馆		推广活动	
	数量(所)	占比(%)	数量(场/次)	占比(%)
前 10% (1-130 所)	9	25	64	42.4
前 30% (131-390 所)	9	25	24	15.9
前 50% (391-650 所)	4	11.1	8	5.3
后 50% (651-1300 所)	8	22.2	29	19.2
无排名	6	16.7	26	17.2
合计	36	100	151	100

注：多家图书馆合作举办的活动各计入 1 次，共举办活动 136 项，总计 151 场/次。

2.3 图书馆强调主场地位和校内合作

开放科学的深入发展不仅要求科研人员积极改变行为范式，也需要科研生态圈中的各个节点构建良好的合作机制以协助落实，与利益相关者密切合作成为高校图书馆开展科研数据管理服务的必要手段之一^[16]。从分析结果来看，本次活动呈现出明显以高校图书馆为核心策划者、组织者和实施者，兄弟图书馆、学科专家、院系、学校机构、社会团体、政府机构、商家等多方参与合作的格局（见表 3）。

从表 3 可见，超过七成活动由单一图书馆承担，这其中有源自图书馆行业的项目，如哈佛大学图书馆的 LibraryCloud 元数据中心推广活动、加州大学图书馆的 HathiTrust 数字图书馆推广活动；此外也有对 Excel、R、Python、Tableau、NetLogo、QGIS、NVivo 等工具的介绍及实际操作，讲解人基本都是数据馆员。这种内容的广泛性体现了高校图书馆对数据服务有较强烈的使命感，强调自身的主场地位，着力凸显在数据驱动型科研中的价值，否则有关工具软件介绍与使用的讲座完全可以委托某些专业人士或团体完成。

观察图书馆联合其他主体发起的活动，可知数据服务是一个多主体协作互动的过程。图书馆与其所在学校的院系及机构合作较多，而与高等教育界以外的机构，尤其是政府机构则合作很少，仅见不列颠哥伦比亚大学图书馆与加拿大统计局合作举办了一次推广使用加拿大住房统计项目（Canadian



Housing Statistics Program) 的讲座。这清晰表明现阶段高校图书馆的着力点还是所在高校。尤其值得注意的是,图书馆与日常业务密切相关的信息产品提供商合作很少,仅见加州大学图书馆与 GALE 公司联合开展的“GALE 数字学术实验室”路演,而该产品是一个商业性平台,仅对付费用户开放,其开放科学属性无疑存在一定的局限性。可见绝大多数图书馆对数据秉承的一致观点是最大程度的“开放、免费”,故有意识地与商业化解方案保持一定的距离。

表 3 高校图书馆合作模式

主办单位	推广活动(项)	占比(%)
单一图书馆	97	71.3
图书馆+院系	17	12.5
图书馆+学校机构	9	6.6
图书馆+图书馆	8	5.9
图书馆+协会	2	1.5
图书馆+社团	1	0.7
图书馆+政府机构	1	0.7
图书馆+商家	1	0.7
合计	136	100

3 活动项目分析

目前学界已提出多种数据生命周期模型,笔者取其中较合适者,将 2022 年“爱数据周”活动分为数个大类以一窥各图书馆的活动焦点。借鉴罗德(Rod A B)等建立的“爱数据周”分析码本^[17]和斯坦福大学博尔吉(Borghi J A)等人归纳的开放科学环境下数据活动项目^[18],将 2022 年活动项目分为六大类,即数据规划、数据管理、数据分析、数据公开、公众科学及数据伦理。

3.1 为项目设计初期的数据规划提供指导

在数据驱动科研的大背景下,主要的科研资助部门对开放数据要求日益增强,这使得科研人员有必要在项目设计阶段就开始规划数据相关任务及时间表,主要包括正确解读政策、规划数据开放路径、选择存储库及进度安排等。美国国立卫生研究院(National Institutes of Health, NIH)要求自 2023 年起,所有资助申请者必须将数据管理及共享方案纳入初始申请书,覆盖面预计涉及 30 万名研究人员和 2500 家机构,以此促进学者将数据管理和共享视为生物医学研究项目的必要组成部分,而不是仅在

规划书的附件材料或声明里表示一下认可^[19]。在此背景下,多家图书馆有针对性地在本次活动中推出了面向项目设计阶段的数据规划辅导,如布朗大学图书馆适时举办了“如何为 NIH 数据管理及共享新政策做准备”的讲座,专门邀请 NIH 科学数据共享政策司司长等政策制定及执行人到场与学者互动;魁北克大学图书馆就如何具体应对加拿大卫生研究院、自然科学和工程研究委员会、社会科学和人文科学研究委员会项目申请的数据政策举办了讲座;利物浦大学图书馆邀请了惠康基金会(Wellcome Trust)介绍其在帮助推广和嵌入开放研究实践方面可发挥的作用;约克大学图书馆举办了数据规划研讨会,内容涵盖数据组织、存储、安全及共享等一系列内容;还有一些图书馆介绍了制定数据规划的具体操作方法,涉及 DMP、DMPTool、DMP Assistant 等工具的使用。

3.2 关注可再现性的数据管理

研究数据管理(Research Data Management, RDM)是对数据及相关研究材料进行存储、组织和描述的过程。从学者个人角度来看,RDM 对数据的长期保存、知识管理和个人成果展示都颇有益处^[20]。从更大视野的学术圈看,开放科学还要进一步强调 RDM 在确保数据可再现性方面的作用,以此促进科研交流和发展。在数据密集型研究中,数据可再现性一直是困扰科研人员的最大难题之一,同一个团队内部,甚至研究者本人都未必能使用既有数据再现研究^[21]。其原因是复杂多样的,有时软件和操作系统的版本差异都会导致研究难以再现^[22],因此 RDM 不仅需要管理详细的材料数据,还有必要考虑对与数据相关的工具、代码、特定软件版本、操作系统等诸多方面的信息进行登记^[23]。

2022 年“爱数据周”活动中,有多项 RDM 推广内容已经超越了个人层面的数据保存、查找、格式化等操作,而着眼于挖掘具群体效益的数据可再现性。例如,利物浦大学图书馆、知山大学图书馆和利物浦约翰摩尔大学图书馆联合举办“开放研究与可再现性”研讨会,邀请了英国可再现性网络委员会(The UK Reproducibility Network, UKRN)的专家与参会者共同探讨英国政府的研发路线图,以及如何在开放研究中实现数据的可再现性;弗吉尼亚大学图书馆举办了关于如何使用 R 对原始数据、分析数据、脚本、元数据、自述文件、项目组织和命名进行管



理的讲座,以提升研究的透明度和可重复性;加州大学图书馆、麦克马斯特大学图书馆均展示了各自的 UCLA Dataverse、McMaster Dataverse 对数据可再现性的支持,让学者了解如何通过分享数据使整个研究团体受益。上述活动无疑是图书馆致力于将开放科学理念映射到 RDM 工作中的有力体现。

3.3 大力推广开源数据分析工具

笔者采用美国国家科学基金会的表述,将用于数据挖掘、清理、分析、可视化等方面的设计语言、程序或软件等视为数据分析工具,它们通常与数据紧密结合,是数据驱动型科研的基础设施^[24]。开源数据分析工具对提升数据的透明度和启发性具有非常积极的作用,秉承开放科学理念的研究团队在实践中可能开发开源工具,并发布开放的数据处理协议、详细分析流程等信息,通过传播新工具、新方法使研

究流程更顺畅^[25]。美国国家科学院有报告显示,开源工具通过完全公开允许任何人出于任何目的的检查、使用、更改和发布源代码,能有效促进开放数据使用,确保代码寿命并减少数据重复处理工作,同时让不同项目、学科的团队合作更加容易,科学家借此可获得更大的合作网络。但现实中的学者并非开源数据分析工具专家,他们常觉得此类事务分散了自己的注意力,并对此感到困惑和畏惧^[26]。这给图书馆推广和支持开发开源数据分析工具提供了机会。在本次活动中,高校图书馆对多种开源数据分析工具进行了大力推广,其中较为典型的推广活动如表 4 所示。这符合当前学术圈面对数据技术和理念时的开放科学倾向,也契合了产业界对开源数据分析工具的日益青睐。

表 4 高校图书馆对开源数据分析工具的推广活动

分析工具	主办单位	活动内容
R	加州大学图书馆、弗吉尼亚大学图书馆、麦吉尔大学图书馆、俄克拉荷马大学图书馆	底层小型基础计算模型、基于 R 的数据社区、CAL-ADAPT 气候数据包在 R 中使用、Jupyter Notebook 开发环境下的 R 应用、ggplot2 使用、dplyr 函数、数据分析脚本编写、与预定义的数据可视化交互来执行查询、开发数据集和分析出版物文本、电子表格处理和数据质量控制、贝叶斯数据分析
Python	加州大学图书馆、弗吉尼亚大学图书馆、麦吉尔大学图书馆	非结构化文本数据的格式化和清理、API 请求、Jupyter Notebook 开发环境下的 Python 应用、数据分析脚本编写
Taguette	加州大学图书馆、维多利亚大学图书馆	电子表格数据清理及采集、定性数据和文本分析
RSpace	哈佛大学图书馆	数据管理
Git、GitHub	加州大学图书馆、密歇根州立大学图书馆	协同工作、版本控制
Gephi	加州大学图书馆、坦普尔大学图书馆	社交网络数据分析和可视化
Leaflet	加州大学图书馆	交互式 Web 地图构建
INOSC	魁北克大学图书馆	在线互动的开放科学社区“国际开放科学和学术界网络”(International Network of Open Science & Scholarship Communities, INOSC)
OpenRefine	维多利亚大学图书馆	数据清理和转换
Vosviewer	魁北克大学图书馆	分析学术期刊引用及语义交叉情况,直观描绘“热点”

3.4 提供可靠的数据公开的最佳实践

如何选择可靠的发布渠道是学者公开数据时需要认真考虑的问题。一方面,公开后的数据面临被不合理使用的风险,如英国曾由于涉及毒品的统计数据被误读而影响了政策制定^[27]。此外,一些第三方机构会大量从互联网上采集公开数据并提供批量下载,而他们的数据管理及理解能力往往低于数据初始公开人,由此也极易误导使用者。故此公开数据需要提供较周全的责任声明,这使得学者们深感

头疼,因为他们并非行政人员或法律专家。另一方面,共享数据在其学术生涯中并没有带来明显的回报,学者们因此也不愿费时费力去整理数据并公开,不少人承认他们有许多从未发表或以其他方式提供给科学界的所谓“暗数据”^[28]。以 NIH 2023 年的数据政策为例,尽管 NIH 有专题网站对学者进行操作指导,并推荐了一些较可靠的解决方案,如其自身开发的一系列存储库,以及一批开放科学项目库,包括哈佛大学开放学术项目 Dataverse、美国航空航天局



开放科学转型计划 Zenodo 等,但各类存储库政策及操作的差异仍给科研人员,尤其是对承担了大部分数据处理工作的年轻学者,带来了不小的困惑^[29]。

由此可见,为了实现可信及低风险的数据公开,数据项目急需参考成熟经验并获得专业人员辅助,以最大限度地减少学者负担。在 2022 年“爱数据周”中图书馆积极推介最佳实践及解决方案,如北卡罗莱纳大学图书馆举办了数据披露风险评估及补救的专题研讨会;维多利亚大学图书馆展示了基于 Dataverse 的数据共享方案;俄克拉荷马大学图书馆介绍了如何使用开源存储库“开放科学框架”(Open Science Framework)进行科研协作和数据存储;哈佛大学图书馆介绍了使用 LibraryCloud 进行元数据存储及开放访问;加州大学图书馆举办了“数据共享知多少”活动,推荐了一批最佳实践案例。

3.5 积极引导受众参与公众科学

在开放科学时代,一种新的科研模式正日益兴起,不同专业的学者乃至科学爱好者广泛合作,以科学的方式系统收集、分析数据,共同参与科研全过程,这就是所谓的公众科学(Citizen Science)。公众科学正日益被政府和行业组织视为应对科研和社会问题的一种新手段而得到大力支持,而开放数据则是支持公众科学的有力保障^[30]。但目前仍有不少学者甚至一些数据中心对公众科学心怀偏见,理由

是对参与公众科学的非专业人员的素养,以及公众科学所产生数据的不信任^[31]。

要改变现状,图书馆的作用不容忽视,在我国,中国科学院已连续 18 年举办“公众科学日”活动,所属的各家文献情报中心始终积极参与。对于高校图书馆而言,普遍接受高等教育的大学生无疑是公众科学最好的参与人,反过来公众科学又能促进学生社团的交流及新技术交流推广,为其未来科研打好基础。高校图书馆可以通过提供数据、空间、信息素养等方面的服务对此予以支持^[32]。在本次“爱数据周”中,多家图书馆组织了带有明显公众科学色彩的活动,其中有代表性的活动见表 5。这些活动内容紧扣社会热点,大量使用开放数据,在激发学生参与热情的同时,也加深了他们对开放科学理念的认同。

3.6 组织探讨数据伦理问题

在开放数据环境下,数据成为人们解析、应对各种社会问题的依据,数据伦理已经与人们的价值观、世界观转型存在明显关联,进而从根本上影响和塑造社会,有学者依据美国图书馆协会的理念,即“图书馆服务于每个人”,提出既然图书馆立志为所在社区的每个人服务,就必须倡导和维护共同的道德伦理,普及数据公正(Data Justice)、数据团结(Data Solidarity)等意识,以有效避免数据创建及使用活动中的许多伦理问题^[33-34]。

表 5 高校图书馆的公众科学推广活动

推广活动	主办单位	活动内容
公众科学:让社区参与数据收集	北卡罗莱纳大学图书馆	邀请学者分享在公众科学项目上的工作经验,详细说明为什么公众科学能够很好满足某些科研需求,以及如何创建项目、与公众沟通、处理数据
监测被遗忘地带的水质:密西西比河杰克逊市案例	布朗大学图书馆	邀请关注社区健康和环境公平的专家介绍一项关于在杰克逊市开展的清洁水项目
数据创建和传播:新冠肺炎疫情期间的教学模式	布朗大学图书馆	演示如何收集学校数据和新冠肺炎疫情数据,教会参与者如何才能创建合适的数据集
新冠肺炎疫情期间被挟持的标签:“我的身体我做主”案例	布朗大学图书馆	多学科学者联合展示对“我的身体我做主”标签含义在新冠肺炎疫情期间的变化及其影响的追踪研究
费城的新冠肺炎与枪支暴力问题	坦普尔大学图书馆	演示一项关于新冠疫情期间费城枪支暴力增多的研究,并探讨内在的关联及社会原因
新冠肺炎疫情数据查找指南	瑞尔森大学图书馆	图书馆员系统介绍从全球到本地的多种新冠肺炎疫情开放数据库
绘制加拿大自行车基础设施地图	西蒙弗雷泽大学图书馆	讲解如何使用加拿大自行车道舒适性和安全分类系统及开放街区地图绘制高质量的自行车基础设施数据,从而推进社会公平和安全研究
使用“微观共享整合数据系列”开放数据创建社会科学数据集	加州大学图书馆	讲解全球最大的人口普查开放数据——微观共享整合数据系列(Integrated Public Use Microdata Series)使用方法

在 2022 年“爱数据周”中,有 9 项活动与社会公正、多样性和包容性问题密切相关(见表 6),另有多

个关于数据安全的活动也在一定程度上探讨了数据伦理问题,主要是关注研究对象的隐私权问题,如布



朗大学图书馆与学校研究诚信办公室、赞助项目办公室联合举办的讲座,向学者介绍如何在国际性合作中避免数据披露与相关法律法规产生冲突;加州大学图书馆的“敏感数据处理”讲座提醒学者在共享数据时要牢记数据伦理。北卡罗莱纳大学图书馆的“数据隐私与在线行为”研讨会旨在让学者了解采集

个人数据如何影响人们在互联网上的行为,以及人们应该如何保护自己。不难看出,图书馆作为信息机构以及社会文化服务体系重要的一分子,在推动数据科学的同时也在积极引导人们进行反思,进而促进数据公平和团结,这是图书馆履行社会责任的重要体现。

表 6 高校图书馆的数据伦理推广活动

推广活动	主办单位	活动内容
数据伦理	塞顿霍尔大学图书馆	数据专家和图书馆员与参与者共同探讨数据伦理问题
数据伦理与公正	加州大学图书馆	一场跨学科的数据伦理与公正研讨会
数据与女性权利	加州大学图书馆	探讨如何将女性权利理论和反思应用于科研及数据处理
《数据偏袒》:与作者对话	布朗大学图书馆	共同探讨大数据和机器学习中由于默认条件设置不当而导致的阶层分裂和种族隔离加剧问题
数据揭秘原住民被奴役史	布朗大学图书馆	介绍“复原美洲原住民被奴役史”项目中登记的 15 世纪至今美洲原住民被奴役数据及实例,探讨他们在殖民地建设中所起的作用
原住民的数据权力	不列颠哥伦比亚大学图书馆	关注加拿大原住民的数据权力,探讨如何使用相关数据
原住民地理空间工具的数据合作及共享经验	不列颠哥伦比亚大学图书馆	分享原住民和其他团体为解决复杂环境、社区和健康问题而开发和使用地理空间门户工具的经验 and 见解
加拿大人口普查局民族、种族及原住民身份数据检索指南	瑞尔森大学图书馆	介绍利用加拿大人口普查局探索民族、种族和反种族主义数据,探讨相关伦理问题的提出及面临的发展
一份数据爱好者的潘多拉文件	麦吉尔大学图书馆	当事记者介绍如何通过向多家离岸服务提供商的数据进行研究,揭露西方政客的贪腐现状

4 实践启示

新时代数字学术的发展有赖于开放科学,数据的开放及有效交流是其中最重要的一环^[35]。在我国,已有部分高校图书馆参与或主导建设了开放数据平台,如北京大学图书馆的“北京大学开放研究数据平台”、武汉大学图书馆的“高校科学数据共享平台”、复旦大学图书馆的“复旦大学社会科学数据平台”等,但总体而言参与主体不多,平台功能及数据开发成效仍有很大提升空间^[36]。有研究指出当前我国高校图书馆建设开放数据平台受到政策、经济、教育及技术等多方面的制约^[37],笔者以为,始终缺乏一项持续的、有较大影响力的推广普及活动也是国内高校图书馆开放数据服务发展较慢的重要原因之一。依据对 2022 年“爱数据周”活动的调研,本文尝试从活动发起者、内容涵盖面、活动项目策划及形式等方面提出若干建议。

4.1 充分发挥图书馆行业组织的带动作用

追溯“爱数据周”的历史,其发展并非一帆风顺。该活动起初由印第安纳大学和普渡大学印第安纳波

里斯联合分校图书馆(Indiana University—Purdue University Indianapolis)于 2016 年的情人节期间开始举办,但在完成 2020 年度活动后,该图书馆决定不再主办并关闭了相关网站,仅保留网页存档,理由是“活动已呈现去中心化发展”^[11],笔者推断其真实含义是活动涉及地区之广、高校之多及事务之繁已令该馆难以负荷。在此关键之时,北美的数据馆员们发起了一次大讨论,最终密歇根大学的高校政治和社会研究联盟(Inter—university Consortium for Political and Social Research, ICPSR)勇担重任,自愿作为新的主办者继续组织活动^[17]。

从这段过往历史可知,行业组织在推广新事物、新服务时,能发挥重要的沟通、协调、统筹能力,其作用是巨大且无可替代的。我国高校图书馆各类工作委员会、联盟等也是馆际交流、统一行动不可或缺的组织,担负着促进高校图书馆整体发展的重任。2014 年由复旦大学图书馆牵头 9 家头部高校图书馆成立的“中国高校研究数据管理推进工作组”,明确将“宣传推动研究数据的科学管理、推荐或提供可



信任的研究数据管理与共享的平台与工具、提供最佳实践案例”等工作列为目标^[38]。笔者认为该工作组极具提升影响力的潜能,可用更开放的态度扩充成员并在图书馆界推出更多活动。目前复旦大学图书馆已联合多家企事业单位举办了 4 届“慧源共享”高校开放数据创新研究大赛,其中的“数据悦读”学术训练营模块就具备大规模推广活动性质,但内容偏重学科专业领域,建议组织更多图书馆参与其中,在国内以合作方式开展一系列基于图书馆的宣讲活动,提升学者及图书馆员的数据素养,在国际上积极开展交流,探索融入合适的数据中心、管理组织和研究机构等的可能性。

4.2 用开放科学的视野拓展活动内容涵盖面

罗德(Rod A B)等通过对往届“爱数据周”活动的分析指出,尽管各参与图书馆都体现了共同的“爱”,但对“数据”的解读却不尽相同,这表明图书馆正日益将“数据”视为一个领域或泛概念并不断扩充其内涵^[17]。2022 年的活动同样符合这一判断。虽然大多数图书馆专注于严格意义上的科研数据工具和方法,但利物浦大学图书馆、知山大学图书馆和利物浦约翰摩尔大学图书馆将“爱数据周”与同期的“开放科学周”(Open Research Week)融合,举办了关于开放获取、开放教育资源的活动,以及坦普尔大学图书馆、塞顿霍尔大学图书馆、蒙特勒阿尔理工学院图书馆举办的关于地理空间数据等活动,都应当被视为跨越了狭义的科研数据,已属开放科学和大数据范畴;此外麦吉尔大学图书馆、北卡罗莱纳大学图书馆、蒙特利尔大学图书馆等还将介绍 Excel 的使用方法也纳入了“爱数据周”推广活动,类似的还有针对 Python、R 等工具的培训等,其实我国高校图书馆已开展不少类似的读者培训,但或许大多数图书馆馆员并未意识到这也是一种对数据和开放科学的支持。笔者认为,图书馆不必过分拘泥,而要充分理解受众的数据行为习惯,既要引导受众“应该”知道什么,又要考虑他们“希望”知道什么,从而策划内涵更丰富的活动,充分调动受众群体的参与积极性。

4.3 设计涉及开放科学多方面内容的“打包”式活动

让更多科研工作者理解并支持开放科学理念是 2022 年“爱数据周”活动的重要任务之一,单纯宣讲开放科学的宏观理念和长远效益恐难引起受众足够的兴趣。本次活动中,有一些图书馆就积极从受众

感兴趣的内容入手,在解决用户切身问题的同时推广开放科学。如利物浦大学图书馆、知山大学图书馆和利物浦约翰摩尔大学图书馆联合推出的“开放科学+”系列活动,涉及科研政策、信息资源、研究方法 & 教学,包括解读基金资助政策的“开放科学与资助者”、指导学者满足科研可再现性要求的“开放科学与可再现性”、交流开放研究实操的“开放科学与我”、交流开放获取图书工作经验的“开放获取与图书”以及交流开放教育资源的“开放教育与国家教学存储库”等。此外,还可以考虑围绕某一项数据实操技能推出系列活动,如麦吉尔大学今年就围绕数据可视化做文章,除一场专题介绍外,还安排了两场关于 R 语言中的 ggplot2 包的使用研讨会,图书馆强调其内容对初学者绝对友好,主要是带领参与者学习使用 ggplot2 创建有吸引力的图形的基础知识,从 tidyverse 入手,导入数据并使用 dplyr 函数进行基本的实操,帮助用户发现和规划自己的数据分析项目。

4.4 用开放活泼的活动形式促进数据科学社区融合

2022 年“爱数据周”活动显示出图书馆对工作坊形式有明显偏好,活动中有 58.2% 是某种形式的工作坊,使得研讨时间、场所、规模、设备等更具弹性;开放平等的氛围和轻松的交流形式,加上主持人及时的沟通修正,使参与者能接触到多种情境下的实操体验,从而更直接地将习得技能转化为科研数据能力。

除工作坊外,数据游戏同样值得关注。已有多项研究表明玩游戏所需的一些技能属于信息素养范畴,采用低压力的游戏模式推广图书馆服务可以有效缓解用户的紧张感^[39-40]。美国图书馆的游戏服务由来已久,开放数据则进一步提供了新题材。2022 年“爱数据周”中有多家图书馆设计了别出心裁的数据游戏,如佐治亚州立大学图书馆的“为图表加标题”游戏,每天发布一张没有标题或数轴标签的图表,邀请师生运用发散性思维为图表加上有趣的标题,图书馆会在社交媒体上分享有创意的标题,并引导学生了解图书馆提供的数据服务和资源;克莱姆森大学图书馆举办了数据隐私风险游戏,参与者要找出数据行为中的风险点,了解如何保护个人隐私,每名参与者都可以得到价值 10 美元的礼品卡;韦尔斯利学院图书馆为数据游戏特制了“爱数据”曲



奇等。在我国,上海图书馆举办的“开放数据竞赛”获奖作品约有 20% 采用了游戏化开发方法,通过将文史数据与游戏场景结合成文化创意产品,更好地发挥了图书馆开放数据的价值^[41]。就高校而言,北京大学、清华大学、武汉大学等大学图书馆都有开发信息素养题材游戏的经验,且他们均是前述“中国高校研究数据管理推进工作组”成员馆,完全有能力开发类似“拯救小布”这样兼具故事化、美术效果及教育功能的数据题材游戏,还可以借鉴上海图书馆的先进经验,在“慧源共享”高校开放数据创新研究大赛中引入公众参与游戏开发,并与有潜力的参赛团队开展合作,开发寓教于乐的数据游戏产品,为高校图书馆数据服务提供宣传新工具。

5 结语

我国《科学数据管理办法》明确提出,要按照“开放为常态、不开放为例外”的共享理念,充分借鉴国内外先进经验和成熟做法,发挥科学数据的重要作用^[42]。利用数据服务为开放科学提供整体支持已成为国外高校图书馆面临的重要挑战和机遇,我国在这方面的需求也日益凸显。尽管“爱数据周”创办时间不长,原组织者对早期活动缺乏系统的数据统计和归纳总结,有针对性的研究目前也仅见罗德(Rod A B)等一文^[17],但考虑到其作为由高校图书馆主导的数据服务密集案例,对“爱数据周”进行进一步调研仍然具有积极的参考价值。故此本文对 2022 年“爱数据周”面向开放科学的数据服务进行梳理,以期为国内同行提供有益借鉴。值得期待的是,ICPSR 将在 2023 年“爱数据周”活动中对活动组织者及参与者进行调研,而且更多世界顶尖大学图书馆,如斯坦福大学图书馆、普林斯顿大学图书馆等,以及我国的香港中文大学图书馆也将作为新组织者参加活动^[43-44],这些变化有望使“爱数据周”更具有行业代表性。

受作者自身研究能力和篇幅所限,本研究仍存在一些局限,如由于各图书馆对“开放科学”和“数据”的理解不同,又或单纯出于推广自身服务的愿望,导致其开展的活动是否真的都完全切合主题或值得商榷;又如 QS 世界大学排名前 10% 的 130 所高校仅有 9 所参加了 2022 年“爱数据周”活动,其代表性尚不够广泛,其余未参加活动的图书馆所做相关工作及其启示,文中则难以涉及;此外研究颗粒度

也有不足,后期将继续有针对性地开展深入研究,以期更好地掌握国际上相关工作进展及可借鉴之处。

参考文献

- 1 Xu C, Yang C. Introduction to big geospatial data research[J]. *Annals of GIS*, 20(4): 227-232.
- 2 Ramjoué C. Towards open science: the vision of the European commission[J]. *Information Services & Use*, 2015, 35(3): 167-170.
- 3 OECD. Making open science a reality[EB/OL]. [2022-03-17]. <http://dx.doi.org/10.1787/5jrs2f963zs1-en>.
- 4 联合国教科文组织. 开放科学建议书草案[EB/OL]. [2022-03-21]. https://unesdoc.unesco.org/ark:/48223/pf0000378841_chi?7=null&queryId=133f0f10-cfbc-478b-a573-42ffdcfb493f.
- 5 Tzanova S. Changes in academic libraries in the era of open science[J]. *Education for Information*, 2020, 36(3): 281-299.
- 6 New York University Libraries. Data services lab[EB/OL]. [2022-04-07]. <https://library.nyu.edu/departments/data-services>.
- 7 Oliver J C, Kollen C, Hickson B, et al. Data science support at the academic library[J]. *Journal of Library Administration*, 2019, 59(3): 241-257.
- 8 University of Connecticut Library. Opening doors to research success: data management programming and outreach[EB/OL]. [2022-04-07]. https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1133&context=acl_nec_conf.
- 9 Purdue University Library. Using health and social data to predict student success: a data competition[EB/OL]. [2022-04-07]. <https://guides.lib.purdue.edu/c.php?g=995885&p=7225864>.
- 10 University of Waterloo. Archives unleashed project scales up with Archive-It[EB/OL]. [2022-04-07]. <https://uwaterloo.ca/arts/news/archives-unleashed-project-scales-archive-it-better>.
- 11 IUPUI. About Love Data Week[EB/OL]. [2022-03-22]. <https://scholarworks.iupui.edu/handle/1805/23785>.
- 12 ICPSR. Love Data Week 2022[EB/OL]. [2022-02-26]. <https://www.icpsr.umich.edu/web/about/cms/3728>.
- 13 夏琬钧, 江艳萍, 赵颖梅. 国外高校图书馆科学数据服务进展与启示[J]. *四川图书馆学报*, 2019(3): 67-70.
- 14 李正超. 国外高校图书馆科学数据服务调研及启示[J]. *图书馆学研究*, 2018(19): 79-84.
- 15 王利君. 英国大学图书馆科研数据服务实践研究[J]. *图书馆界*, 2019(6): 43-47.
- 16 Federer L. Research data management in the age of big data: roles and opportunities for librarians[J]. *Information Services & Use*, 2016, 36(1): 35-43.
- 17 Rod A B, Isusterb M Y, Chandler M. Love Data Week in the time of COVID-19: a content analysis of Love Data Week 2021 events[J]. *The Journal of Academic Librarianship*, 2021, 47(6): 102449.
- 18 Borghi J A, Gulick A. Promoting open science through research data management[EB/OL]. [2022-06-22]. <https://doi.org/10.48550/arXiv.2110.00888>.
- 19 NIH. Final NIH policy for data management and sharing[EB/OL]. [2022-07-05]. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>.
- 20 杨鹤林. 数据监护: 美国高校图书馆的新探索[J]. *大学图书馆学报*, 2011, 29(2): 18-21.



- 21 Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility[J]. Proceedings of the National Academy of Sciences, 2018, 115(11): 2584–2589.
- 22 Gronenschild E H, Habets P, Jacobs H, et al. The effects of free surfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements[J]. PLOS ONE, 2012, 7(6): e38234.
- 23 Grüning B, Chilton J, Köster J, et al. Practical computational reproducibility in the life sciences[J]. Cell Systems, 2018, 6(6): 631–635.
- 24 NIH. Final report of national science foundation advisory committee for cyber infrastructure task force on software for science and engineering[EB/OL]. [2022-07-24]. https://www.nsf.gov/cise/oac/taskforces/TaskForceReport_Software.pdf.
- 25 Lowndes J S, Best B D, Scarborough C, et al. Our path to better science in less time using open data science tools[J]. Nature Ecology & Evolution, 2017, 1(6): 1–7.
- 26 National Academies of Sciences, Engineering, and Medicine. Open science by design: realizing a vision for 21st century research[EB/OL]. [2022-06-12]. <https://doi.org/10.17226/25116>.
- 27 Stimson G V, Hickman M, Turnbull P J. Statistics on misuse of drugs have been misused[J]. British Medical Journal, 1998, 317(7169): 1388.
- 28 Heidorn P. Shedding light on the dark data in the long tail of science[J]. Library Trends, 2008(2): 280–299.
- 29 Kozlov M. NIH issues a seismic mandate: share data publicly[J]. Nature, 2022, 602(2): 558–559.
- 30 Fecher B, Friesike S. Open science: one term, five schools of thought[EB/OL]. [2022-08-20]. https://doi.org/10.1007/978-3-319-00026-8_2.
- 31 Shanley L A, Parker A, Schade S, et al. Policy perspectives on citizen science and crowdsourcing: citizen science: theory and practice[EB/OL]. [2022-05-20]. <https://doi.org/10.5334/cestp.293>.
- 32 苏幼婷, 许春漫. 美国高校图书馆公众科学服务实践及启示[J]. 图书情报工作, 2019, 63(3): 129–137.
- 33 Syracuse University School of Information Studies. Social justice and librarianship[EB/OL]. [2022-04-17]. <https://ischool.syr.edu/social-justice-librarianship>.
- 34 Braun M, Hummel P. Data justice and data solidarity[J]. Patterns, 2022, 3(3): 100427.
- 35 The Royal Society. Science as an open enterprise[EB/OL]. [2023-02-23]. <https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf>.
- 36 李洋, 温亮明. 我国高校图书馆科学数据开发现状调研与分析——以一流大学建设高校图书馆为例[J]. 图书馆工作与研究, 2021(12): 5–15.
- 37 韩滢莹. 开放数据背景下中国高校图书馆数字学术服务环境分析[D]. 武汉: 武汉大学, 2018.
- 38 上海市图书馆学会. 中国高校研究数据管理推进工作组成立[EB/OL]. [2022-04-17]. <http://society.library.sh.cn/node/2611>.
- 39 Leer L V. Interactive gaming vs library tutorials for information literacy: a resource guide[J]. Indiana Libraries, 2006, 25(4): 52–55.
- 40 Giles K. No budget, no experience, no problem: creating a library orientation game for freshman engineering majors [J]. The Journal of Academic Librarianship, 2015, 41(2): 170–177.
- 41 谭雅文. 图书馆开放数据的游戏化开发路径探析[J]. 农业图书情报学报, 2022, 34(6): 83–92.
- 42 中国政府网. 让科学数据开放共享成为常态[EB/OL]. [2022-08-27]. http://www.gov.cn/zhengce/2018-04/05/content_5279957.htm.
- 43 The Chinese University of Hong Kong Library. International Love Data Week 2023[EB/OL]. [2022-08-27]. <https://www.lib.cuhk.edu.hk/en/about/events/love-data-week-2023/>.
- 44 ICPSR. International Love Data Week 2023 events[EB/OL]. [2022-08-27]. https://www.icpsr.umich.edu/web/about/cms/3799?utm_source=all&utm_medium=all&utm_campaign=LDW23&utm_id=LDW23.

作者单位: 暨南大学图书馆, 广东广州, 510632

收稿日期: 2022 年 8 月 31 日

修回日期: 2023 年 2 月 27 日

(责任编辑: 关志英)

Accelerating Open Science: A Content Analysis of Academic Libraries Events in Love Data Week 2022

Yang Helin

Abstract: The open science movement continues to gain momentum, attention, and discussion. As science research has become increasingly data-driven, data services now play a critical role in enabling and accelerating open science to academic libraries. Through case study, content analysis and literature review, this study analyzes the elements and characteristics of 136 academic libraries events in Love Data Week 2022. Some suggestions in programs organizer, contents and activity ways are presented for libraries hosting data services and events.

Keywords: Academic Libraries; Open Science; Open Data; Love Data Week; Data Service