

北京爱迪科森教育科技有限公司协办

学术图书馆的科研数据管理服务

吴新年 (中国科学院兰州文献情报中心 兰州 730000)

摘要 本文在系统调研国内外学术图书馆科研数据管理服务实践的基础上,分析总结了目前国内外学术图书馆在科研数据管理服务方面开展的主要工作,归纳提出了学术图书馆在开展科研数据管理服务过程中需要注意的几个问题,以期对计划开展科研数据管理服务的学术图书馆提供参考。

关键词 学术图书馆 科研数据管理 数据服务

On the Scientific Data Management Services of Academic Libraries

Wu Xinnian (Lanzhou Documentation and Information Center, CAS, Lanzhou, 730000)

Abstract In this paper, the author analyzes kinds of research data management services in academic library, based on investigation of practices on research data management service in academic libraries all the world, and summarizes some problems need to pay close attention to in the process of extending research data management services, hopes to provide the reference for these academic library which plan to carry out the research data management services.

Keywords academic library, research data management, data service

1 引言

人类社会正处在一个“大数据”时代。2013年被媒体称为“大数据元年”,它标志着我们的生活和思维方式因大数据再次发生潜移默化的改变,世界正在开启一次重大的时代转型:一个大规模生产、分享及应用数据的时代^[1]。然而,据2011年美国Mckinsey Global Institute发布的Big Data: The Next Frontier for Innovation, Competition and Productivity调查报告显示,尽管全球数据量飞速增长,但仍有将近87.5%的数据未得到真正利用,许多数据资源并没有形成真正的知识源以供研究人员利用^[1]。而随着科学研究范式的变化和数字科研模式的发展,尤其是科学界对开放数据、数据管理规划的需求,以及对潜在“大数据”研究机会的日趋重视,学术机构越来越需要能够提供一系列更为便捷完善的数据服务,以满足研究过程中各个阶段的特定要求。面对这种日益强盛的新型需求,许多学术图书馆积极响应,纷纷采取行动,不断拓展服务内容,开拓新型服务模式,探索开展科研数据管理服务。例如,哈佛大学图书馆已经引入“大数据”服务^[3];美国普渡大学图书馆建立了分布式数据管理中心D2C2并开始

提供服务^[4];哥伦比亚大学图书馆研究制订了详细的研究数据存储实施战略,指导和推进科研数据的管理服务^[5]。还有人美国研究图书馆协会(ARL)成员馆进行了调研^[6],结果表明,约93%的成员馆至少提供帮助科研人员查找并使用科学数据资源的服务,约84%的成员馆还同时提供地理空间分析服务,约79%的成员馆同时提供数据集获取服务,约74%的成员馆同时提供版权与专利咨询服务;大约50%的成员馆计划近期新增一项或更多服务,特别是数据可视化、数据挖掘和数据分析服务。同样在这次调查中发现,约75%的成员馆开展了科研数据管理服务,内容包括:参与制订数据管理计划或提供相关咨询与指导、科研过程中的数据管理指导(例如对数据存储或文件安全提出建议)、研究文件与元数据、结题项目或已公开发布的科研数据共享与管理(如遴选、保存、存档、引用等);另外有23%的成员馆计划在不久的将来提供上述服务。由此可见,科研数据管理服务已经成为学术图书馆未来业务的重要内容之一。

2 学术图书馆科研数据管理服务的主要内容

根据当前国内外学术图书馆,尤其是美国ARL成

员馆的实践探索,学术图书馆在科研数据管理服务方面可以开展的服务工作将主要包括五个方面。

2.1 参与科研数据管理计划(DMP)的制订与实施

目前,美国许多学术图书馆都提供这类服务,尤其是美国国家科学基金会(NSF)、美国国立卫生研究院(NIH)等研究资助机构明确提出所有项目申请者在提出项目研究计划申请的同时必须同步提供关于未来科研数据管理的规范计划要求后,许多学术图书馆都积极采取行动,有的为科研数据管理计划的制订提供咨询与指导,有的直接参与科研数据管理计划的编制,还有的甚至会继续跟进,参与相关计划的实施。例如绝大多数的美国 ARL 成员馆(约 87%)提供有关科研数据管理计划的在线资源,包括对 NSF、NIH 等科研资助机构的数据管理计划要求与指南进行详细解释,帮助研究人员更好地理解相关政策与制度,以便研究制订出符合规范要求的数据管理计划;还有的成员馆在线提供创建数据管理计划所需的工具(如 DMP-Tool¹⁾)或资源以及 DMP 样板范例、有关数字仓储服务、长期数据管理与保存、其他相关资源链接、版权信息以及相关培训信息等。这些在线资源有些是自建的,也有的是定制第三方的(如 DMPTool)。图书馆在相关计划的后续执行中也是积极参与,主动承担数据的整理、组织、管理与服务工作以及提供相关的设施设备和环境条件支持。

2.2 DMP 准备工作以外的科研数据管理服务(RDMS)

除了根据有关政策规定帮助研究人员制订数据管理计划之外,也有许多学术图书馆还普遍提供数据管理最佳实践案例(内容涉及文件格式与命名规范、数据存储与备份、任务对象与 IRB、知识产权(IP)与版权、元数据、存档与保存、引用与发表数据、版本控制以及其他相关在线资源与研讨会等),帮助科研人员确定并应用适当的元数据标准,规范地完成研究文件组织与命名、数据引用、数据共享与获取、数据存储与备份、数据安全等。已经有越来越多的学术图书馆提供关于科研数据管理的咨询与培训服务。这些咨询与培训,有的是图书馆工作人员独自完成的,有时候图书馆也只是发挥桥梁、纽带或服务平台的作用,借助于图书馆所联系的数据专家来帮助一般研究人员解决关于科研数据管理的问题。至于咨询与培训的方式,多数是通过电子邮件、即时通讯软件或电话形式完成的,也有许多是在科研人员的办公室、实验室或其他地点以会谈形式完成的。而实践证明,不同的形式,效果是不一样的,多是根据当次要讨论问题的复杂程度确定。由于当面讨论与培训能够最大限度地深入交流,因而最能够保证咨询与培训的效果。还有的学术图书馆在探索提供数据出版、数据权限管理以及数字图像数据转换等服务。

2.3 数据存档服务

目前在许多国家,对于公共资金资助的科学研究数据都要求开放共享,已经成为公益服务的一部分。在这种背景下,作为长期提供公益性基础知识服务的图书馆似乎自然而然地就成为提供或参与此类服务的当然候选者。在美国等一些发达国家,许多学术图书馆正在尝试开展此类服务,而科研人员也普遍愿意到图书馆寻求有关数据存档与传播方面的帮助,例如美国 ARL 成员馆中约 74% 的成员馆都在提供数据存档服务²⁾。

至于数据存档服务的具体实现形式,不同的图书馆所采取的方式和途径是不同的:有的是建设专门的数据档案库——有专门的硬件设备与软件系统用于科研数据的管理与服务;而更多的是借用已有的机构知识库(IR),即通过拓展 IR 的功能,以支持对科研数据的管理与服务。从目前的情况看来,这两种方案似乎在一定程度上都能够满足当前对科研数据管理与服务的需求,但长远发展下去哪一种效果会更好,还暂时无法做出准确的评判。倾向性的意见是专门的数据档案库有利于长远发展需求,但由于其中涉及大量资源与经费的投入,对于图书馆到底能有多大实力维持和发展这种服务仍然是必须考虑的问题。也正是因为这样,绝大多数学术图书馆目前暂时选择依托 IR 实现科研数据管理与服务,因为对于大多数学术图书馆来说,拓展服务的需求是明显增长的,但经费与关键资源投入并没有同步增长,许多新增服务的成本必须由图书馆自己去解决。当然也有的学术图书馆获得了相关基金会等机构提供的数据管理服务专项经费投入,但这毕竟还是极个别的,对于科研数据管理服务的经费投入并没有成为所有学术图书馆经常性经费的一部分被正常考虑在内。

2.4 科研数据关联集成服务

孤立的数据是没有太大意义的,只有相互关联起来,才可能挖掘出数据的更多内涵与意义。

首先是科研数据与科学文献的关联集成。图书馆是存储科学文献的机构,对于科学文献的管理与传播服务既是学术图书馆的本职工作,也是其最擅长的业务。随着馆藏文献资源的数字化,目前许多图书馆已经能够实现多元资源的关联化组织和主动推荐型服务。因此,在科研数据成为学术图书馆管理与提供服务的资源之一后,图书馆工作者必然会想到将科研数据与其他馆藏资源,尤其是科学文献的关联化组织。例如,将某一数据或数据集与依托这一数据或数据集产生的科学文献关联起来,从而使得以后有用户检索到某篇科学文献时,能够很轻易地探查到该文献依据的数据或数据集;同样的,当用户检索到某一数据或数据集的时候,也会很轻易地发现该数据支持过什么样的研究,支持产生过什么样的科学文献,从而帮助用户更好地认识数据和理解数据,更进一步用好数

据。还有的学术图书馆通过将科学文献中的数据挖掘出来,形成各类专题的数据集,并以文献为纽带,将这种二次数据或再生数据集与原始数据集建立起关联,方便后来的研究者选择利用。当然,这里所谓的科学文献也不仅仅是指期刊论文、会议论文或学位论文,还可能是专利、标准、科技报告、学术专著、教材或讲义、课件等。

其次是数据之间的关联集成。尤其是多学科海量数据汇聚与集成,支持实现数据资源的跨领域、跨学科的关联集成组织,以促进学科的交叉和多学科协同创新研究。

此外,还有科研数据与其他多元资源的关联集成,包括与机构信息、人员信息、科研项目、会议信息、设施设备、模型、图像、三维地图等的关联集成,构建全源信息网络,优化研究信息环境。

2.5 其他相关服务

随着在科研数据领域工作的不断深入,一些学术图书馆也正在积极参与数据政策制定、数据组织标准规范建设。还有的学术图书馆探索了协助开展专题数据挖掘、数据挖掘与分析工具开发、多元数据质量与可信度验证、大规模多元数据的动态交互式可视分析、数据回溯存储、数据再标引、数据资源组织与数据检索系统建设、数字数据长期保存系统建设、灾备恢复等服务。

3 学术图书馆开展科研数据管理服务过程中需要注意的几个问题

科研数据管理服务是学术图书馆一项新兴的业务,也是学术图书馆适应当前及未来大数据、数字科研和数据密集型科研的必然选择。然而,在发展这种新型服务过程中,有些问题是必须要引起足够关注与重视的。

3.1 经费投入问题

在目前图书馆经费普遍紧张的情况下,自身能否有效地解决拓展科研数据管理服务所需的成本问题,是所有拟开展科研数据管理与服务的学术图书馆必须首先要考虑清楚的问题。虽然学术图书馆的经费一直随着社会的发展在增加,但与其发展需求相比,一直存在较大的差距。而新兴的科研数据管理服务与传统业务相比,所需要增加的成本要更大,比如更大型的硬件设备与设施,更复杂的软件系统与工具等,这些都需要大量的经费投入,而且有些投入还必须是长期持续不断的。虽然目前已经开展科研数据管理服务的学术图书馆有些有幸得到额外的专项经费支持,但大多数学术图书馆还是要依靠自身能力解决开拓新服务的成本。因此,学术图书馆能否争取从其原经费渠道增加可用于科研数据管理服务的经费,或者获得额外专项支持科研数据管理服务的稳定经费,是决

定学术图书馆能否顺利、持续地做好这一新型业务的关键。

3.2 数据存储系统平台的选择问题

目前许多学术图书馆选择使用开源软件作为其全部或部分科研数据存档方案,也有一些图书馆开发了自有软件。如前所述,许多学术图书馆是在 IR 中拓展了科研数据管理服务,因此 DSpace 就成为这类学术图书馆普遍使用的系统平台;也有部分学术图书馆选择使用 Fedora 作为其 IR 平台,因而配合使用其他相关工具(如 Hydra 或 iRODS)实现科研数据管理服务;也有的图书馆使用了 Fedora 和 Data Conservancy 平台的软件 Chronopolis,这是 HubZero 的一个定制版本,是包含 Fedora、Archivematica、Dataverse、iRODS 和自定义构建的知识库在内的综合数字资源系统。至于选择什么样的系统平台,应该根据拟管理的数据类型及服务需求来决定。尤其是致力于长久发展科研数据管理服务的学术图书馆,应该考虑更专业的科研数据管理系统平台,并注重系统平台的可升级、可扩展性和可交互性以及系统平台的安全性问题。

3.3 科研数据的采集途径与方式

IR 也是近年来学术图书馆发展起来的一项新业务,在 IR 发展过程中一直令建设者头痛的事情就是内容的采集或收集问题。目前大多数学术图书馆建设的 IR 中存储着期刊论文、会议论文、学位论文、研究报告、电子课件、演示文稿等,这些类型的资源除了由图书馆工作人员通过一些手段集中采集和导入之外,正在通过各种鼓励或强制政策督促科研人员自助提交。好在 IR 系统平台正越来越友好,通过不断增加智能化辅助工具尽可能地减少科研人员的提交成本(包括能力成本和时间成本等),来鼓动和维持科研人员的自助提交热情。但是,与学术论文、研究报告、课件和 PPT 文稿相比,科研数据要复杂得多,其管理元素要多得多,提交过程中的人机交互也多,能不能保证科研人员有足够的耐心和热情,是保证未来科研数据管理系统数据数量和数据质量的关键。

3.4 数据安全控制问题

首先是数据信息安全问题。一般地,学术图书馆的传统资源是不太涉及保密问题的,因而其使用往往是不受限制的、充分开放的。而 IR、RDMS 虽然也是开放的,但有些资源和数据有可能是有限开放的,或者部分资源和数据对有些用户是有限开放的,只有获得充分授权的用户才有可能无障碍使用所有的资源和数据。例如,纽约大学图书馆参与建成的网络视频数据图书馆 Databrary 的全部访问权限就只提供获得授权的研究人员^[9]。因此,必须事先解决好对数据的信息安全控制问题,不但要事先明确政策机制,还应该解决好技术实现问题。

其次是数据的有效定位问题,即数据的永久标识

符问题。目前几乎所有的数据档案馆都建立了唯一标识符制度,从而解决了数据或数据集的长期定位问题,并使其在出版物中得到可靠引用。例如在 IR 中最为普遍的是 Handle System^[9],而大多数数据档案馆则使用 DOI 来标识数据集,也有少数数据档案馆使用 ARK^[10],还有一些数据档案馆生成了自己的标识符。

第三是数据的长期有效性维护问题。为了始终保持良好的数据管理服务,有的学术图书馆还定期进行数据文件完整性检查,提供多重副本复制,定期进行数据文件格式校验或转换升级,这也是保证数据安全的重要方面。

此外,与数据安全有关的还有数据版权、数据隐私、数据伦理等问题,也应该事先有所考虑并及时完善相关的政策与机制。

3.5 人员配置与发展模式的重构问题

数据管理服务是学术图书馆的一项新兴服务,由于该业务从内容到方法都与传统的文献服务、信息服务业务有较大不同,因而对管理服务人员的类型及素质的要求也大不相同。目前已经开展科研数据管理服务的学术图书馆中,许多馆都设立有专门的数据馆员岗位,招聘有专门的人员专职从事科研数据管理服务^[11-13];有的图书馆还设置了专门的服务部门提供数据管理服务(如普渡大学图书馆^[14]);有的图书馆是成立了专门的数据管理服务工作组专责提供科研数据管理服务^[2];也有的是通过培训提升图书馆已有人员,尤其是学科馆员或联络馆员的数据管理能力,使其能够兼职承担起科研数据管理服务的工作;有的学术图书馆则招聘数据专家或有数据管理专长的研究生业余兼职在图书馆协助解决有关科研数据管理服务的难题;还有的学术图书馆是通过跨界合作,加强与专门的数据档案馆、数据中心、科研管理部门、计算中心、出版机构等的合作,实现对科研数据管理服务的介入。例如,美国俄勒冈州立大学图书馆与该校出版社签订了科研数据服务战略协议,共同建构支撑数据驱动型科研基础设施^[21]。

也正由于这是一项新兴的业务,一切都还在探索之中,因而大家的经验和体会都是不同的,尤其是关于从事科研数据管理服务的人员应该具备什么样的素质与技能,似乎也还没有一个公认的统一评价标准。但是较普遍的看法是,最好是拥有某一学科的专业背景,然后又掌握和能够熟练运用情报学、计算机科学的方法与技术的人员。因而一些学术图书馆倾向于要求相关工作人员能够具有某一专业学科和信息科学(情报学或计算机科学)的双学历和学位。

随着新型业务对人员素质和专业技能的新要求,图书馆工作人员需要从技术和思维上做好双重准备去应对这种变化^[2],尤其是要有数据思维和大数据意识。现代社会,数据意识很重要,它包含着精确与开

放、透明与分享等因素,相信数据、用数据来说话,是理性精神的一种表现^[23]。而为了满足大数据时代对于数据管理人才的需求,美国一些高校或数据管理机构已经尝试开设了有关数据管理的课程,并逐步发展起一门新兴学科——数据科学,旨在培养专业的数据管理专家。还有更多的则是通过不定期地举办短期培训班以及提供在线课程学习,帮助相关人员(包括数据科学家、数据管理者、数据分析师、学科馆员、档案馆员等)提高能力。加拿大研究图书馆协会(CARL)还设立了数据管理分会这样的专门组织,开设数据管理服务网上课程,提供在线培训,推进数据管理及数据馆员职业发展^[24]。此外,一些项目组也会提供一些免费学习资料,如 RDMRose 研究小组提供的在线学习课程^[25]。

4 结语

人类的科学研究活动在经历了实验科学、理论科学、计算仿真科学的时代后,目前已经迎来了数据密集驱动型科学的时代。随着科学研究越来越协作化、数据密集化和计算化,科研人员日益面临大量的数据管理服务需求。学术图书馆作为支撑科学研究的重要公益性机构,积极参与,部分或全部地承担起所在机构的科研数据管理服务任务,构建适应和支撑数据驱动型科研的基础设施,将是学术图书馆适应大数据社会的必然选择。尽管在这一发展过程中学术图书馆将面临诸多的困难和挑战,但也面临许多新的发展机遇。能否抓住这些发展机遇,突破自身的能力陷阱,实现创新性发展,对于当前每一个学术图书馆来说都将是一次战略性抉择。

参考文献

- [1] 张学军. 大数据视野下图书馆的新走向[J]. 山东图书馆学刊, 2013(6): 32-36.
- [2] McKinsey Global Institute. Big Data: The Next Frontier for Innovation, Competition and Productivity [EB/OL]. [2014-03-21]. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.
- [3] LISNews. Harvard Releases Big Data for Books [EB/OL]. [2014-03-21]. http://lisnews.org/harvard_releases_big_data_for_books.
- [4] 洪程. 国外科学数据服务现状研究[J]. 图书馆杂志, 2012(10): 31-34.
- [5] Columbia's Evolving Research Data Storage Strategy [EB/OL]. [2014-01-25]. <http://academiccommons.columbia.edu/catalog/ac:162637>.
- [6] SPEC Kit 334: Research Data Management Services [July, 2013] [EB/OL]. [2014-01-25]. <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>.
- [7] DMPTool: Guidance and Resources for Your Data Management Platform [EB/OL]. [2014-01-25]. <https://dmp.cdlib.org/>.
- [8] Scientists Undertake Effort to Launch Video Data-Sharing Library for Developmental Science [EB/OL]. [2014-03-15].

- <http://www.nyu.edu/about/news-publications/news/2013/07/02/scientists-undertake-effort-to-launch-video-data-sharing-library-for-developmental-science.html>.
- [9] Handle System[EB/OL]. [2014-04-15]. <http://www.paradigm.ac.uk/workbook/metadata/pids-handle.html>.
- [10] Archival Resource Key(ARK)[EB/OL]. [2014-04-15]. <http://www.paradigm.ac.uk/workbook/metadata/pids-ark.html>.
- [11] 孟祥保,钱鹏.国外高校图书馆数据馆员岗位设置与管理机制[J].图书与情报,2013(4):12-17.
- [12] Scientific Data Curator[EB/OL]. [2014-04-03]. <http://library.brown.edu/about/employment.php>.
- [13] Data Librarian(Librarian II or III)[EB/OL]. [2014-04-03]. <http://www.iassistdata.org/resources/jobs/1647>.
- [14] Data Librarian[EB/OL]. [2014-04-03]. http://www.brocku.ca/hr/careers/position_detail.php?id=897.
- [15] University of Edinburgh Job Description[EB/OL]. [2014-04-03]. http://www.docs.csg.ed.ac.uk/HumanResources/JDs/8-Data_Librarian_EDINA_ISG.pdf.
- [16] Research Data Librarian, Hewa 7, QUT Library[EB/OL]. [2014-04-03]. http://www.and.s.org.au/guides/dmframework/data/librarians_andproject_qut.pdf.
- [17] Data Librarian[EB/OL]. [2014-04-03]. <http://www.iassistdata.org/resources/jobs/3009>.
- [18] Research Data Curator[EB/OL]. [2014-04-03]. <http://www.iassistdata.org/resources/jobs/3170>.
- [19] Data Services Specialist[EB/OL]. [2014-04-03]. http://ils.indiana.edu/careers/view_job_specific.php?job_id=13578.
- [20] Cornell Research Data Management Service[EB/OL]. [2014-02-25]. <https://confluence.cornell.edu/display/rdmweb/services>.
- [21] Oregon State University Libraries & Press Strategic Agenda for Research Data Service[EB/OL]. [2014-02-21]. http://ir.library.oregonstate.edu/xmlui/bitstream/handle/1957/38794/OSULP_StrategicAgendaForResearchDataServices_2013-2014.pdf?sequence=4.
- [22] [英]伊安·约翰逊.智慧城市、智慧图书馆与智慧图书馆员[J].陈旭炎,译.图书馆杂志,2013(1):4-7.
- [23] 聂细文.大数据时代面临的统计挑战[N].中国信息报,2013-01-08(7).
- [24] CARL Course: Introduction to Research Data Management Service[EB/OL]. [2014-02-25]. <http://www.carl-abrc.ca/en/scholarly-communications/data-management-sub-committee/introduction-to-research-data-management-services.html>.
- [25] RDMRose Learning Material[EB/OL]. [2014-02-25]. http://rdmrose.group.shef.ac.uk/?page_id=10.
- [作者简介] 吴新年,男,1968年生,中国科学院兰州文献情报中心研究员。
- 收稿日期:2014-04-18

欢迎订阅 2015 年

图书、情报、信息、资料工作者 自己的刊物

《情报资料工作》

CSSCI 来源期刊,全国中文核心期刊,中国人文社会科学核心期刊
双月刊,大 16 开,112 页,全年定价 288 元,国内统一刊号 CN 11-1448/G3

——图书馆、情报室、资料室、信息中心的理想文献、必藏刊物

——图书、情报、信息、资料、档案工作者的业务参谋、良师益友

《情报资料工作》杂志面向图书情报界、信息产业界和文献资料档案界,读者遍及高校、党校、科学院、军队院校、政府信息机构及公共图书馆系统。刊物一贯注重追求理论精品,面向工作实际,形成了求实创新的学术风格,是广大图书情报工作者进行学术研讨及业务交流的重要园地。

订阅方式:

(1) 在全国各地邮局订阅,邮发代号:82-22

(2) 直接向中国人民大学书报资料中心市场部订阅

① 邮政汇款:

地 址:北京 9666 信箱(邮编 100086)

收款人:市场部

电 话:010-82503412/38/40 62512171

② 银行汇款:

户 名:中国人民大学书报资料中心

开户行:中国银行北京人大支行

账 号:344156031742

欢迎致电《情报资料工作》编辑部,电话:010-62512296/62514977;或登录博客:<http://blog.sina.com.cn/qbzlgz>;登录网站:<http://qbzl.zlzx.org/CN/volumn/home.shtml>;或发邮件至 qingbaoziliao@263.net。