

从开放获取到开放知识环境

——高校图书馆数字馆藏发展政策研究与实践

孙晓菲, 张婷, 陈蕾

(浙江大学图书馆, 杭州 310027)

摘要: 从环境、政策和评价体系等方面分析开放获取的现状和发展趋势, 探讨基于数据的科学研究第四范式对数字馆藏发展政策的影响, 提出面向开放知识环境的数字馆藏发展核心政策和保障政策。

关键词: 开放获取; 开放知识环境; 馆藏发展政策

中图分类号: G250.76

DOI: 10.3772/j.issn.1673-2286.2015.12.008

开放获取 (Open Access, 简称OA) 源自布达佩斯开放存取倡议 (BOAI), 是指科研人员将论文、专著、图书、演示稿、数据、课件等研究成果发表在开放式学术期刊或存放于开放式知识库中, 以免费的、不受限的方式供读者阅读、下载、保存和利用。

开放获取的理念由来已久, 但开放获取运动如火如荼地开展是在互联网时代, 其本质是数字时代学术交流模式的转型和发展。纸媒时代, 大部分学术期刊集中在少数商业出版机构手中, 出版商控制学术出版领域, 加速学术出版商业化进程, 形成日益严重的市场垄断, 学术交流渠道狭窄、滞后、备受控制。随着互联网和信息技术的发展, 学术交流不再以传统的纸媒模式为主, 而是依托网络环境, 这使得知识交流的内容、形式和渠道趋于多样化, 知识传播呈现时滞短、效率高的特点, 交流范围覆盖整个互联网。

开放的学术交流模式推进了开放数据和开放教育, 构建了全新的开放知识环境。面对开放带来的学术环境的颠覆性变化, 高校数字图书馆建设应及时调整建设思路, 制定前瞻性的数字馆藏发展政策, 以适应学校教学、科研与管理在开放知识环境中的新需求。

1 开放获取现状和发展趋势

传统的学术交流分为正式交流与非正式交流两种

基本类型。正式交流是指经过同行评议的学术成果, 以期刊论文、会议论文或著作的形式传递给其他学者。非正式交流则是直接发生于学者之间的交流, 其内容包括理论知识、问题假设、研究程序和方法等^[1]。网络时代, 信息的数字化和信息网络的建立, 成熟的OA期刊、广泛的数据分享及平等的教育理念为学术交流方式的改变奠定了基础, 逐步改变图书馆被动接受由出版商主导的学术交流体系。开放获取不仅赢得了学术界和图书馆的支持, 还引起了各国政府的高度重视, 从政府层面, 制定了强制开放获取政策, 以遏制学术垄断。同时, 为了提高OA期刊的质量, 保证学术的严肃性与可持续发展, 面向OA期刊的评价体系也在不断努力建设中。

1.1 三足鼎立的开放获取环境

学术期刊作为学术资源的重要组成部分, 实现开放获取能打破出版商对学术资源的垄断, 加强学术信息的交流, 提高学术交流的时效性和信息使用的直接性和交互性, 提升科学研究的公共利用程度, 实现资源共享和公平使用, 减少用户之间因为费用引起的信息鸿沟。因而在近十几年, 无论数量、来源范围还是影响力, 都得到了迅速发展。以全球最大的OA期刊门户DOAJ为例, 其已收录的OA刊来自136个国家, 出版机

构既有Springer、Wiley、Nature、Elsevier等出版商，也有IEEE、APS、AMS、AGU等学会和协会组织，还有牛津大学、剑桥大学、中科院、印度科学院等为代表的研究机构^[2]。2006年4月底，DOAJ仅收录2,200余种OA刊，95,400余篇OA论文^[3]；而截至2015年3月底已收录10,351种OA刊，1,859,430篇OA论文^[2]。同时，开放获取期刊的影响力也逐渐增强，例如作为独立学术出版机构的BioMed Central（简称BMC）目前出版277种同行评议的OA期刊（2015年3月），其中175种刊被JCR收录^[4]。总之，开放获取期刊迅速增长的发展趋势以及不断提升的影响力，使之逐步逼近“成为学术研究主流资源”的转折点^[5]。

开放数据的分享随着科研的发展也日益为公众所关注。特别是在2007年1月11日，美国微软研究院计算机科学家吉姆·格雷博士在对计算机科学和远程通讯委员会作的演讲报告中，提出了科学研究范式在经历了第一范式经验（Empirical）范式、第二范式理论（Theoretical）范式、第三范式计算（Computational）范式后，现已发展到第四范式（the Fourth Paradigm），即数据密集型科研范式（Data-Intensive Paradigm）^[6]。开放科学数据体现了“信息共享和开放科研”这一现代科研理念，正在推动科研范式的变革，重塑科学交流模式。

除了学术期刊和研究数据，教育资源也是重要的科研成果产出。2001年，美国麻省理工学院（Massachusetts Institute of Technology，简称MIT）率先拉开了OER建设的序幕，宣布“将其所有的课程资料在线发布，使每个人都能访问到”^[7]。近年来，开放课程运动已从镜像传统课堂发展到大规模在线开放课程（Massive Open Online Courses，简称MOOC）。MOOC不仅突破传统课程时间、课堂空间以及课程人数的限制，能够满足大规模课程学习者学习，而且整合了多种形式的数字资源形成丰富的课程资源，利用多种网络社交工具形成多元化的学习工具，使教育过程可以真正成为沉浸式、参与式、交互式、合作化和自适应的过程，以配合不同学习者的目标。

1.2 强制OA政策和OA评价体系

开放获取已获政府的政策支持，现共有40多个国家颁布了强制性OA政策。OA仓储强制归档政策注册网（The Registry of Open Access Repositories

Mandatory Archiving Policies，简称ROARMAP）^[8]为大学、研究机构以及研究基金提供注册，以记录强制OA政策的发展历程，要求研究人员通过开放式访问提供有同行评审的研究成果。在该注册网上，科技发展水平较高的欧美国家都制定了政府OA支持政策，如欧洲的Horizon 2020项目^[9]、美国白宫科技政策办公室（US White House Office of Science & Technology Policy，简称OSTP）^[10]以及国家科学基金会（US National Science Foundation，简称NSF）^[11]的OA政策、英国的研究理事会及联合政府政策（RCUK&Coalition Government Policy）^[12]等。这些政策或从经费上支持OA，或是从学术论文开放的时限上支持OA。

此外，为了解决开放资源的评价问题，出现了不同于传统评价的网络文献计量学（Altmetrics）以及ORCID组织针对学术规范而建立的研究者信用评价体系。这两者对OA学术评价都具有深远意义，并引起了广泛关注。

Altmetrics一词由北卡罗来纳州大学教堂山分校的图书情报博士生Priem Jason于2010年在其Twitter上首次提出。Altmetric指标主要是用来衡量网络环境下学术研究的影响力，其最初目的是提供文章级论文影响力测度指标（Article-Level Metrics）。与传统期刊影响因子等引用指标相比，其特征与潜在优势^[13]包括评价信息源更广、评价维度更多、评价粒度更小、评估反馈更及时等，突破了现有文献计量研究理论，支持一切非传统研究成果纳入评估体系，将有利于提高跨学科的信息传播速度。

ORCID（Open Researcher and Contributor ID，开放研究者与贡献者身份）则是通过为研究者配置唯一的并可链接到其研究成果的身份标识码，为搜索提供规范统一的研究人员以及著作成果信息，确保研究者的作品信用，减少确定研究人员学术成果的时间，利于学者与相关资助者、出版商和机构跟踪其科研项目，查找个人学术论文更容易、更精确。从发展前景看，ORCID嵌入科研与学术研究生态系统是趋势所在。

2 面向开放知识环境的图书馆数字馆藏发展政策研究

图书馆虽然已在数字环境中经过二十多年的历练和发展，但仍偏重基于文献资源的集成服务模式，与传统服务模式同出一辙。于是，图书馆不得不面对这样的

现实,由商业数据库商及出版社主导的文献资源供应体系将图书馆绝大多数信息资源局限于传统的书刊资源,而基于描述性元数据的资源整合与发现系统,缺乏语义搜索与知识发现服务功能。目前,基于集成服务的数字图书馆提供的馆藏数字信息资源集成检索与获取面临着迅速边缘化的严峻挑战^[14],而数据库商再次走到图书馆的前面,他们提供的第三方知识服务产品在开放信息环境或用户知识环境中,支持用户进行科学信息内容的分析、可视化,并建立个人机构知识库,管理各类知识资源。因此,图书馆要建立来自第三方的文献资源分析、管理与挖掘平台,受到数据库商的各种限制。

要改变被动局面,开放合作是图书馆的最佳选择,联盟式开放知识环境能使参与各方享受到大数据时代共性与个性的最优资源获取与资源服务。对数字图书馆来说,外部业务的构建与内部业务一样重要,如何构建开放知识环境,制定开创性的数字馆藏发展政策将成为数字图书馆的研究重点。

2.1 科学研究第四范式对图书馆数字馆藏发展政策的影响

“范式”研究是美国科学史家托马斯·库恩1962年在《科学革命的结构》(*The Structure of Scientific Revolutions*)一书中首次引入科学史研究的。他认为,“任何一个科学领域在发展中达到成熟的标志”,是必须“取得了一个范式,取得了范式所容许的那类更深层的研究”^[15]。

库恩的范式论是关于科学发展的一般方法论,一经提出,就引起了哲学家和科学家的关注和热烈讨论,以及社会科学家的积极响应,各学科专家都纷纷运用范式论来分析和说明各自学科的历史发展过程,图书馆也不例外。在格雷的E-science演讲中,他特别提到在第四范式影响下的新型数字图书馆概念。他认为:新型图书馆内容应该是免费的,但管理内容和同行评审却不便宜;我们不仅需要出版物的档案库,更需要数据的档案库。报告的最后,他建议:培育数字化的数据图书馆,建设既支持数据也支持文献的数字图书馆^[16]。

格雷的提议对数字图书馆来说,并不是件轻松的事。他的数字图书馆概念里,其实包涵三个层面的建设任务:第一层面是图书馆的数字图书馆;第二层面是新型的数据图书馆,是针对第四范式的科学研究所产生的大数据而建立的,是具有图书馆意义的数据库;第三

层面是后数字时代的图书馆,是“数字图书馆+数据图书馆”,一种新型的开放学术知识空间或环境的建设。从文献到信息、从信息到知识,图书馆的主要保存对象除出版物为主体的文献资源外,还将包括以开放信息为主体的数据资源和以知识服务为主体的增值资源。

以上分析可知,第四范式对数字图书馆的影响是革命性的,它的目标不是改造现有的数字图书馆,而是创建一个新的数据图书馆,与现有的数字图书馆共同构成未来的数字图书馆。

2.2 开放知识环境下数字馆藏发展政策的战略重点

面向开放知识环境的数字馆藏发展政策研究,不是孤立的数字馆藏研究,它不仅涵盖静态的传统文献资源,而且涵盖动态的现代开放资源。由于内涵与外延的拓展,数字馆藏发展政策不仅要建立起传统与现代的桥梁,而且要兼顾到静态与动态的差异,以适应图书馆在新技术环境中的发展要求。开放合作、跨界融合、协同管理、标准规范和永久保存五个方面将是数字馆藏发展政策的战略重点。

其中,开放合作和跨界融合是数字馆藏发展政策的重要核心战略。面对大数据,图书馆除了资源合作,还应扩展到知识组织与管理的合作、知识服务的合作及知识发现的合作,更多地解决海量灰色资源及科学数据的发现、组织、管理与利用。通过开放知识空间的合作,将图书馆转化为支持合作学习、合作研究的新型公共知识平台。由于开放知识环境不受文献类型、阅读对象、收藏功能、社会功能等限制,具有最广泛、最复杂的数据来源,因此传统意义的图书馆、档案馆、博物馆等公共基础文化机构在数字知识空间将逐渐合为一体,文献、档案、文物等数字资源将登陆同一平台,融合并形成荟萃人类文化遗产、世界档案、历史文献以及先进科技成果的超级知识空间。

而协同管理、标准规范以及永久保存是数字馆藏发展政策的重要保障战略。协同管理政策将为开放知识环境中的参与方提供与数字资源相关的元数据管理、数据共建共享、资源检索与发现以及学术知识创新等公共服务政策。标准规范政策提供整个建设体系的发展政策与框架结构,规定标准规范的建立与维护机制、标准规范采用的现行标准与规范集、标准规范的文本格式与书写规则、标准规范的执行范围等。永久保

存政策则保证图书馆的纸本文献、电子文献以及开放科研与开放教育等各种开放资源通过协同管理政策,达到永久保存的目标,以节约单个数字图书馆的保存成本,提高资源的保障率与利用率,充分发挥开放知识环境的协同优势。

3 开放学术知识环境的建设与实践

对高校图书馆来说,开放知识环境除了外部的OA期刊、网络公开课程等教研资源外,高校本身作为知识传播与研究的重要机构,从教学、科研到管理都积累了大量的数据资源,大部分可以在校园内开放获取,对了解学校的发展动态具有较高的研究价值。随着智慧校园的建设,图书馆、教研机构、管理机构等都希望通过合理的模式来解决各类资源特别是开放学术知识资源长期保存和统一揭示的问题。

浙江大学近年来在知识资源信息服务方面作了许多有益的尝试,从网络学术创新环境到机构知识库的建设,为学校教学科研资源的保存利用积累了大量的资源储备和技术储备,打造全校性的一体化知识资源采集、组织、管理和利用的信息化服务体系,提供面向互联网各种固定和移动终端的知识资源的跨平台应用,通过用户对资源的利用情况进行大数据分析,建设和谐的开放学术知识环境生态系统。

浙江大学的开放知识环境整体规划以知识资源中心为核心,辐射教学与科研,形成“一体两翼”的框架模式,如图1所示。通过资源中心的标准化接口,可以汇聚从系统采集、跨系统交互及用户主动性提交等多种途径获取的知识资源。这些资源可以是资源实体,也可以是资源元数据,或者是资源相关性标签等。资源中心经过对知识资源的聚合、关联等方法,将知识资源提供给各应用系统或者学术科研平台、网络教学服务平台。用户可以通过各种平台、客户端以及移动终端,获得各种知识资源服务。

知识资源中心作为学校知识云平台的重要核心组件,为一系列知识云服务提供基础。在设计思路,知识资源中心提倡开放互联和开放式应用,满足各种开放资源的采集、存储与利用需求。在开放互联上,知识资源中心提供开放式资源接口,各应用系统或开发者可以利用数据接口对知识资源进行二次开发,以向用户提供多样性的知识资源服务。知识资源中心提供基于传统互联网和移动互联网模式的跨平台支持。在开放式应用上,知识资源中心需要对资源分类不断丰富与完

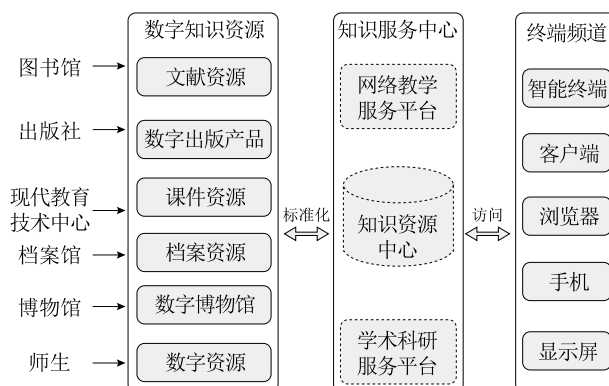


图1 知识资源中心总体框架图

善,制定科学的数据规则、灵活的功能服务,实现与外界系统的关联,服务调用方式灵活、简单、有效。知识资源中心以用户体验为根本,充分利用已有项目的技术成果积累,形成开放式集成的同时,保障系统的高效、安全、稳定。

4 结语

开放获取作为互联时代新的知识传播与学术交流方式,为世界提供了开放存取知识的机会,其发展结果将改变现有的科学出版性质与机制,改变现有的科学评价体系。开放获取支持下的学术交流模式淡化了学术出版的商品属性,恢复了学术成果的公共物品性质。随着开放获取的发展,图书馆的馆藏范式将发生根本性变化。图书馆不再满足于文献的收集,或是将数据作为附件加在文献末尾,而是致力于各种学术数据的采集和学术过程服务。数字图书馆除了文献资源外,将为从数据获取(Data Capture)、数据监护(Data Curation)、数据分析(Data Analysis)到数据可视化呈现(Data Visualization)的研究全过程,提供数据支持服务,实现数据采集、索引、存储、检索、提取以及与出版文献的互联,完成数据的重构与知识的创新。

从开放获取到开放知识环境建设,数字馆藏发展政策将成为图书馆馆藏发展政策的核心,引领图书馆跨入数据互联时代。

参考文献

- [1] 崔慧仙.网络时代的学术交流——以科学网为场景的观察与研究[D].上海:华东师范大学,2011.
- [2] DOAJ[EB/OL].[2015-03-25].<http://doaj.org/>.

- [3] 刘辉.开放获取期刊数据库——DOAJ[J].四川图书馆学报,2006(5):30-34.
- [4] BioMed Central[EB/OL].[2015-03-26].<http://www.biomedcentral.com/about/faq/impactfactor>.
- [5] 张晓林,李麟,刘细文,曾燕.开放获取学术信息资源: 逼近“主流化”转折点[J].图书情报工作,2012, 56(9):42-47.
- [6] Hey T, Tansley S,Tolle K. Jim Gray on eScience: A Transformed Scientific Method[C]//The Fourth Paradigm—Data-Intensive Scientific Discovery.USA: Microsoft Research,2009:17-31.
- [7] About OCW[EB/OL].[2015-03-26].<http://ocw.mit.edu/about/>.
- [8] Welcome to ROARMAP[EB/OL].[2015-03-26].<http://roarmap.eprints.org/>.
- [9] The Eu Framework Programme for Research and Innovation HORIZON 2020 [EB/OL].[2015-03-26]. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- [10] US White House Office of Science & Technology Policy (OSTP) [EB/OL].[2015-03-27]. <http://roarmap.eprints.org/529/>.
- [11] US National Science Foundation (NSF) [EB/OL].[2015-03-27]. <http://roarmap.eprints.org/528/>.
- [12] RCUK Policy on Open Access.[EB/OL].[2015-03-27].<http://www.rcuk.ac.uk/research/openaccess/>.
- [13] Piwowar, H. Introduction altmetrics: What, why and where? [J]. Bulletin of the American Society for Information Science and Technology,2013,39(4):8-9.
- [14] 孙坦.开放信息环境:学术图书馆信息资源建设的重定义与再造[J].中国图书馆学报,2013(5):9-17.
- [15] 库恩.科学革命的结构(第四版)[M].北京:北京大学出版社,2012:17,169.
- [16] Tony Hey,Stewart Tansley,Kristin Tolle.第四范式:数据密集型科学发现[M].潘教峰,张晓林等,译.北京:科学出版社,2012:5,11,23.

作者简介

孙晓菲, 浙江大学图书馆资源建设部主任, 副研究馆员。

张婷, 浙江大学图书馆资源建设部馆员。

陈蕾, 浙江大学图书馆资源建设部馆员。

From Open Access to Open Knowledge Environment: A Study for Digital Collection Development Policy and Practice of University Libraries

SUN XiaoFei, ZHANG Ting, CHEN Lei
(Zhejiang University Library, Hangzhou 310027, China)

Abstract: The article analyses the open access development status based on the facts of open access resources, government open access policies and open access review systems. It discusses that the forth paradigm of science study has influence over the digital collection development policy. It also proposes the core polices and guaranty policies of the digital collection development policy to open knowledge oriented environment.

Keywords: Open Access; Open Knowledge Environment; Digital Collection Development Policy

(收稿日期: 2015-12-09)