

国外开放科技信息资源建设情况分析 & 启示

王星¹, 苏静², 杨代庆¹

(1. 中国科学技术信息研究所, 北京 100038; 2. 武汉大学信息管理学院, 武汉 430072)

摘要: 开放信息环境下, 开放科技信息资源已成为学术交流生态系统的关键环节, 是一项必不可少的现代科学生产要素。选取国外科技领域开放期刊、开放图书、开放会议资源、开放学位论文、开放科学数据等五类信息资源, 论述近年来开放科技信息资源的建设现状及建设特点。基于此, 探讨国家层面对开放资源统一管理、开放资源协同长期保存机制构建、多元开放平台内容衔接关联的现实必要性。

关键词: 开放获取; 科技信息资源; 信息资源建设; 开放资源评价

中图分类号: G203

DOI: 10.3772/j.issn.1673—2286.2014.11.002

1 引言

2001年12月布达佩斯会议召开以来, 学术界对开放获取的概念和组织形式等问题逐步达成共识, 并通过一系列宣言明确了开放获取的内涵和外延。经过十余年发展, 传统科学知识交流方式发生巨大变化, 开放信息环境极大推动着开放科技信息资源朝向多元化、规模化、集成化方向发展。在资源内部建设方面, 可开放获取的科技文献, 即通过OA期刊发表的、提交至机构知识库自存储的以及可直接在网络获取的科技类学术论文, 已约占全部科技期刊论文的20%, 并预估该数字会呈现逐年上升趋势^[1]。在资源外部使用方面, 开放科技信息资源可向更广泛的用户群体提供信息服务, 尤其是为企业用户创建了便利条件。英国联合信息系统委员会 (Joint Information Systems Committee, JISC) 的调查报告则显示, 对于广大中小企业来说, 开放获取资源尤其重要, 因为中小企业普遍缺少科技资源需求的预算, 并认为在当前中小企业竞争压力不断加剧的外部环境下, 开放科技信息资源的免费使用可显著节约企业成本^[2]。此外, 有超过90%的研究人员表示在图书馆找不到目标资源后, 会采用互联网搜索免费的科技信息资源^[3]。由此可见, 以“开放、免费”为特性的科技信息资源

已成为学术交流生态系统的关键环节, 是一项必不可少的现代科学生产要素。在世界范围内公开科学运动影响下, 我国强制或鼓励开放出版的政策措施积极出台, 开放科技期刊、开放科学数据、开放科技报告、机构知识库、开放资源集成平台等资源建设活动取得阶段性进展^[4-5]。目前国内相关研究主要集中于某一资源类型、建设项目的发展现状评析^[6-7], 或是从国外出版商、基金资助机构、高等院校、国家政府等单一主体来阐述开放获取运动的实践措施^[8-10], 缺乏较为系统反映国外科技类开放信息资源建设全貌的文献论述。基于此, 本文通过对不同类型开放科技信息资源的建设现状进行分析, 管窥当前开放科技信息资源的建设框架、发展特征和成功经验, 并梳理和归纳其在发展进程中的重要问题, 旨在为我国开放科技资源建设实践提供借鉴参考。

2 国外开放科技信息资源建设现状

当前, 开放科技信息资源涉及学术期刊、会议文献、学位论文、图书、研究报告、科学数据、音像等众多类型, 分布于开放获取仓储、数字出版内容投送平台、资源集成目录、社交网络、公共信息发布系统等, 本文

重点从资源类型视角对开放科技信息资源建设现状进行调研分析, 具体探究其建设规模、分布特征、质量评估标准、组织揭示层级、知识共享方式等。

2.1 开放期刊

开放期刊是开放科技信息资源的重要组成部分, 也是开放信息环境影响下发展规模最为突出、参与群体最为丰富、受众影响最为深远的资源类型。本研究通过对DOAJ、BMC、PMC、HighWire、Open J-Gate等若干核心开放期刊平台以及Web of Science、Scopus等重要商业数据库中收录的开放期刊进行实证调研, 统一归并查重后, 对学科不符、停止出版、被列入Beall黑名单等问题数据予以剔除, 最终确定科技类开放期刊7001种。开放科技期刊的来源分布较为分散, 7001种期刊分布于2741个网站, 其中仅有57个网站所含期刊多于10种, 只含1种期刊的网站高达2372个。依据学科类别划分, 医学类开放获取期刊占据总量的41%, 理学、工学、农学的期刊数量则较为平均, 均维持在1400种左右。同时, 7001种开放期刊中明确说明经同行评审的开放期刊为5250种(75%), 至少被1个重要全文或文摘数据库收录的期刊为5040种。主要数据库的开放科技期刊收录情况如表1所示。

表1 7001种开放科技期刊数据库收录情况 (截至2014年8月)

数据库名称	学科领域	期刊数量 (种)
Web of Science	科技综合	1002
BIOSIS Previews	生命科学	505
Zoological Record	动物生物学	530
Medline	生物医学	470
CABI	农林业	1341
INSPEC	物理学、电子工程、电子学、 计算机科学及信息技术	318
FSTA	食品科学和技术	71
Scopus	科技综合	2182

此外, 绝大部分可通过网站抓取、FTP下载和开放接口3种获取全文, 并存在“先注册、后下载”或部分全文可免费获取的情况。约2600种(40%)开放期刊对使用遵循协议进行了明确说明, 多数遵循CC协议, 但是版本略有不同, 其他使用说明还包括SpringerOpen

copyright and license agreement、Taiwan Association for Aerosol Research (TAAR) 和英国政府发布的non-commercial license for public sector information等。

2.2 开放图书

开放图书是开放获取运动中生长周期较短的资源类型。通过调研DOAB、eScholarship、Intech、OpenEdition Ebooks等大型开放图书集成平台, 结果显示: 科技类开放图书特别强调同行评议的严谨性、数字化管理的规范性和资源内容的可发现性, 因此总体呈现出质量标准高、可获取数据格式较统一、外部链接支持度高的资源特征, 大部分遵循CC协议和OAI-PMH元数据收割协议, 便于在线批量获取和集成, 可基本满足资源创建者版权保护权益。截至2014年8月底, 依托大型开放图书出版项目可直接获取pdf格式的开放科技图书已达4000余册。其中, DOAB提供科技图书349本, 涉及领域有科学概论、健康科学、技术工程、地球与环境科学等; eScholarship收录科技图书238本, 其中经同行评议图书106本, 作者拥有所有版权并可选择性使用CC许可, 用户可免费阅读网页版和下载PDF版, 需付费使用POD版; Intech提供科技图书2769本, 分别涵盖物理科学与工程、健康科学、生命科学等, 支持个人用户按章节或整本书在线阅读、下载, 图书馆用户可将PDF文件保存到本地、打印或获取复本供个人使用, 教师可用于教学、会议、讲座等非商业用途, 上述行为均无需注册或付费^[11]。OpenEdition Ebooks主要涉及人文社科、环境及健康领域, 提供科技图书319本, 图书质量均经过技术与学术评估, 以HTML格式出版, 普通用户可通过Web在线免费浏览, 成员图书馆用户可下载定制的PDF和EPUB版资源。

2.3 开放会议资源

开放会议资源包括会议论文、视频、音频、演示文档、日程目录等形态资源, 具有动态性、分散性、易逝性等固有特征, 主要以HTML、PDF格式呈现, 但整体描述层级较为简单、使用遵循协议标识较不明晰。目前, 著名的综合性开放会议文献仓储有涉及多学科的WASET (World Academy of Science, Engineering and Technology)^[12]、计算机系统的USENIX^[13]、计算机语言领域的ACL Anthology (Association for

Computational Linguistics)^[14]等,权威的专业性开放会议数据库有关于空间太赫兹信息技术的ISSTT Proceedings (International Symposium on Space Terahertz Technology)、MIT组织的信息质量会议集ICIQ (International Conference on Information Quality)、美国声学协会主办的POMA (Proceedings of Meetings on Acoustics)、国际光学工程学会出版的SPIE Digital Library等。其中,核心科技类开放会议约有5000余个。具体而言,WASET会议论文均经过同行评审,高质量论文将出版并可获取全文及元数据,其他论文只能获取题名和作者信息;ICIQ要求用户需填写姓名、单位和邮箱信息后才可免费获取原文;SSTT Proceedings不仅可供用户获取全文,还提供一次性批量下载功能。值得说明的是,BMC Proceedings、Elsevier Procedia、IOP Publishing等正式出版的开放会议资源具有较高的存储稳定性和可利用价值。

2.4 开放学位论文

鉴于创建主体及其知识产权保护细则的丰富多样,开放学位论文使用权益要求较为复杂,部分版权声明较为隐蔽,并且元数据格式不完整情况较为普遍,若为广泛应用需加大后期人工成本。其中,eScholarship支持OAI元数据收割协议,遵循CC协议,可获取文摘及全文近1万篇;德克萨斯数字图书馆学位论文数据库可获取文摘及全文超过2万篇,但分为无限制下载、有限制下载和不能下载三种情况,需层层点击、转换页面才能确认最终是否可以下载;ProQuest、NDLTD学位论文库均可提供文摘和部分全文,但文摘数据不完整,全文使用限制条件较多,如全文下载需提前注册、存在无效链接等情况。

2.5 开放科学数据

随着数据密集型科研活动的日益兴起,开放科学数据的发布、使用与权益保护机制的积极构建逐步受到人们的关注。开放科学数据涉及主题领域较多,如多媒体数据、地理数据、社交网络数据、文献出版物数据、多学科实验数据等,提交方式包括独立出版、借助学术论文出版、作为辅助材料出版等。尽管存在异质结构存储与利用、多维数据集成与更新、语义一致性、数

据归属权保护等冲突与问题,开放科学数据的生长活力与交互作用还在不断增强。英国研究委员会RCUK出台规定,要求自2013年4月1日起,接受RCUK资助的科研项目,必须开放获取其成果(金色或绿色OA),实行CC-BY协议,同时还要提交一份相关科学数据访问的说明。阿根廷参议院也在2013年通过有关OA义务化法案,由公共资金所产生的科学技术论文、学位论文要在机构知识库上予以公开,同时强制规定所有研究数据5年后必须公开,且能够被其他研究人员利用^[15]。此外,自然出版集团于2014年5月推出的在线出版开放获取期刊*Scientific Data*,明确要求作者需对数据集进行详细描述,并且将同行评议的重点聚焦于描述的完整性、社群标准的一致性、数据存储的恰当性、数据价值的复用性和测量的技术质量。2014年7月ORCID与DataCite签署科学数据合作备忘录,旨在改善和协调科学数据的存储、发布与共享条件,构建可发现、可引用、互操作强的数据支撑平台^[16]。

3 当前开放科技信息资源建设特点

3.1 开放主体日益多元化,利益关系错综复杂

开放科技信息资源的建设主体多样,主要包括政府部门、学协会组织、高等院校、科研院所、图书馆、出版商等开放政策制定与实践机构。其中,政府部门是保证开放科技信息资源稳定、持续建设的主导力量。随着各国对开放获取的认识不断加深,无论是发达国家还是发展中国家均已开始从国家层面介入并制定相应政策,通过国家立法的方式强制本国利用公共资金取得的科研成果能够公开获取。在此影响下,截至2014年10月,全球137家科研资助机构中接近80%已经出台强制开放呈缴或鼓励开放呈交的政策措施^[17]。此外,出版商之间开始建立战略合作联盟,如2013年初自然出版集团收购瑞士OA出版商Frontiers的控股权,波兰出版集团和加州大学出版社宣布于2015年将期刊资源发布至斯坦福大学HighWire开放平台^[18-19],出版商与国家之间逐步形成订阅优惠机制,如英国物理学会出版社对英国、奥地利提供免费、低价格的期刊采购方式。作为传统信息服务机构,图书馆承担着开放论文处理费谈判和管理的工作职能,有研究表明,超过三分之一的机构为管理开放论文处理费建立了集中式部门,且这类部门中的三分之二由大学图书馆负责^[20]。

3.2 开放数量持续扩大, 资源类型存在差异性

近年来, 伴随全球科学出版业信息公开模式的日臻成熟、数字环境下科研用户协同工作的行为转换, 以及各国政府、资助机构、图书馆等多元主体的政策推动等因素, 开放科技信息资源数量呈现不断增长的发展态势。据汤森路透Web of Science的数据显示, 2013年约有11%的科研论文发布在开放获取期刊, 而在2002年仅有2%^[21]。此外, 2014年7月至9月, DOAB的科技类图书仓储量由104本增加至349本, 增长率为235.6%。在资源整体规模不断增长的同时, 不同类型的开放科技信息资源在学科分布、权益使用、质量管理等方面具有一定差异性。以开放论文为例, 生物医学领域的开放期刊于2011年出版12万余篇科技论文, 占当年开放期刊发文总量的35.5%, 其次为社会科学与人文艺术、地球与环境科学、工程学、物理与天文学等领域^[22]。相比之下, 开放图书出版项目主要集中在人文社会科学领域。此外, 相较于部分开放会议资源, 开放学位论文的知识资产管理意识更为明显、使用权益更为规范。

3.3 开放深度不断拓展, 增值服务效能显著提升

在海量科技信息资源加速扩充的前提下, 如何有效节约读者获取目标资源的时间成本、如何集成组织多元异构来源资源以优化页面展示空间, 如何面向科研用户构建知识交流与知识增值服务, 将是实现开放科技信息资源的深度发现与高效利用的重要课题, 而细化资源揭示粒度和开发资源利用层次是解决上述问题的基础条件。目前, 开放科技信息资源的开放揭示深度呈现出由单一文献类型向跨文献类型整合、由单个机构向跨机构整合、由传统目次与文摘揭示向语义组织整合的发展态势。以生物医学和生命科学领域顶级开放获取期刊eLife为例, 该刊已将论文加工级别深入到图表、科学数据和参考文献, 其中, 前期已被赋予唯一标识码的科学数据可作为图表的附件(Additional files)上传, 以供读者复用关联资源, 图表的显示效果采用鼠标滑动左边小图、右边即时显示大图的网页设计原理, 最大程度地紧凑知识结构且减少版面尺寸。同时, 期刊作者可对自己已发表的论文添加补充解释性说明和延续性研究成果, 直接面向科研用户提供了内容可扩展、可交互的论文增值服务。此外, IPython notebook多媒体资源融合

工具、Authorea科研论文合作系统等也为碎片化、多角度揭示与关联开放资源提供了实用的技术条件。

3.4 开放权限愈加繁杂, 代表性资源共享水平趋于一致

依据资源类型、建设主体、出版年代等因素的不同, 开放科技资源的开放权限和范围各异, 有的支持复合选项、论文处理费折扣或免费的补偿模式(如英国皇家化学学会推出的“Gold for Gold”), 有的遵循CC-BY使用协议、版权方自定义使用协议或是未明确声明的使用协议, 用户须根据规定合理出版和使用开放资源。并且, 伴随着学术交流体系的变革, 科研机构、出版机构、研究人员等主体已广泛认识到提升资源共享水平、最大限度开放科技资源的重要性: 大部分情况下, 大型开放科技信息资源服务平台主要是以CC协议为主; 截至2013年, 开放获取学术出版联盟(Open Access Scholarly Publishers Association, OASPA)内遵循CC-BY协议的会员论文已达12万余篇, 相较上一年度增长4万余篇, 并整体呈现较快的向上态势^[23]; 2014年10月20日起, 自然出版集团将旗下期刊*Nature Communications*由过渡型的复合期刊直接转换至全开放期刊, 不再提供传统订阅模式出版选项。

4 结论

目前, 国外开放科技信息资源的建设主体多样, 资源类型与数量愈加丰富、资源加工揭示级别逐步深入、资源开放力度明显加大, 客观反映其建设发展势头依然强劲。但为实现开放科技信息资源建设的可持续发展, 尤其是有助于我国开放获取政策制定和相关建设项目实施, 应从以下方面加以考虑:

一是强化国家层面对开放资源的统一管理。目前, 政府、科研院所、高等院校、出版机构、学协会组织、图书馆、科研人员等多元主体已成为开放科技信息资源建设的政策主导者和实践开拓者, 但多方利益诉求和价值期盼各有不同。例如, 以盈利性为首要目标的商业出版机构, 必然会以投入与回报比衡量开放获取运动的经济价值, 并以此为依据决策相关政策取向。尽管自然出版集团、Wiley等商业出版商已取得阶段性盈利, 但开放出版的收入仍只占据当年科技期刊出版业总收入的很小比例^[24], 间接影响了开放获取政策的支持力度和

实践积极性。为了能够持续、稳定地面向全球科研人员提供长效的开放科技信息资源保障服务,有必要从国家层面协调、平衡各方利益关系,大力激发多元建设主体的内在驱动力,强化开放激励政策的落实,强调开放源的互操作,并统一建设国家级资源集成开放平台,确保资源公开的透明度和流程管理的规范化。

二是强化开放资源的协同长期保存机制。开放科技信息资源存在易逝性特征。根据西班牙赛博实验室的研究显示,欧洲研究型门户的开放学术资源都存在半衰期,且每种类型资源的半衰期不同,其中网页的半衰期最短,平均只有458天,图片的半衰期有778天,其他应用型资源的半衰期(包括PDF文档)介于二者之间。为避免动态资源的价值流失,有必要对开放科技信息资源的学术价值进行评估,注重资源质量控制与管理,选择性开展高价值开放资源的保护与集成建设工作,从遴选、采集、加工到仓储建设、平台服务等方面,采取阶段渐进式的方式积极引导协同长期保存机制的构建与完善。

三是强化开放平台的内容衔接与关联。开放科技信息资源种类繁多、来源复杂、分布不均衡且内容耦合关联程度松散,有必要开发开放平台的内容活力,运用语义技术优化知识表示、知识获取、知识检索、知识推理等服务功能,在异构环境下解决机构知识库、学科资料库、出版商资源库等平台之间的互操作问题,以增加内容紧密度和可见度。在此基础上,依托开放科技信息资源与商业资源的联合建设、本地馆藏与在线资源的高度融合,形成真正的“大”数据仓储,积极引导科技资源的优化配置和高效利用。

参考文献

- [1] BJÖRK B C, LAAKSO M, WELLING P, et al. Anatomy of green open access [J]. Journal of the Association for Information Science and Technology, 2014, 65(2): 237-250.
- [2] DAVID P, DICK W, JANE H. Benefits to the Private Sector of Open Access to Higher Education and Scholarly Research [EB/OL]. [2014-10-15]. http://open-access.org.uk/wp-content/uploads/2011/10/OAIG_Benefits_OA_PrivateSector.pdf.
- [3] CAROL T, RACHEL V. UK Scholarly Reading and the Value of Library Resources [EB/OL]. [2014-10-15]. <http://www.jisc-collections.ac.uk/Reports/ukscholarlyreadingreport/>.
- [4] 张晓丹,王莉,梁冰.开放获取在NSTL系统中的实现与服务模式研究[J].数字图书馆论坛,2013(10):58-62.
- [5] 张晓林,刘细文,李麟.研究图书馆推进开放获取的战略与实践:以国家科学图书馆为例[J].图书情报工作,2013,57(1):15-19,48.
- [6] 陈雪飞,黄金霞,张建勇,等.国外开放科技资源项目实证研究及启示[J].图书馆建设,2013(6):18-25.
- [7] 张春玲,朱江.国外开放教育资源(OER)建设情况分析 with 启示[J].图书馆论坛,2013,33(4):120-125.
- [8] 杨小薇, MARGO B.哥廷根大学开放获取发展战略与举措[J].图书情报工作,2014,58(7):53-58.
- [9] 李国俊,邱小花,肖明.国外高校强制性开放获取政策研究[J].大学图书馆学报,2014(3):39-43.
- [10] 孙博阳,李书宁.国外大学图书馆开放获取实践概述[J].图书情报工作,2013,57(10):18-24.
- [11] Linking terms and conditions [EB/OL]. [2014-10-15]. <http://www.intechopen.com/js/ckeditor/kcfinder/upload/files/linking%20terms%20and%20conditions.pdf>.
- [12] Conferences by Research Areas [EB/OL]. [2014-10-15]. <http://waset.org/Conferences>.
- [13] Conferences USENIX [EB/OL]. [2014-10-15]. <http://www.usenix.org/publications/library/proceedings/>.
- [14] A Digital Archive of Research Papers in Computational Linguistics [EB/OL]. [2014-10-15]. <http://www.aclweb.org/anthology/index.html>.
- [15] Argentina Passes Open Access Act For Publicly Funded Research [EB/OL]. [2014-10-15]. <http://www.ip-watch.org/2013/12/16/argentina-passes-open-access-act-making-publicly-funded-researchavailable/>.
- [16] ORCID and DataCite sign an MoU [EB/OL]. [2014-10-15]. <http://www.datacite.org/node/131>.
- [17] Some JULIET Statistics [EB/OL]. [2014-10-15]. <http://www.sherpa.ac.uk/juliet/stats.php?la=en& mode=simple>.
- [18] Portland Press Ltd to move to the HighWire Open Platform [EB/OL]. [2014-10-17]. http://highwire.stanford.edu/PR/PortlandPress_HighWire.pdf.
- [19] University of California Press to partner with HighWire [EB/OL]. [2014-10-17]. <http://www.ucpress.edu/blog/16448/university-ofcalifornia-press-to-partner-with-highwire/>.
- [20] JISC Collections Open Access Fees Project: final report [EB/OL]. [2014-10-17]. <http://www.jisc-collections.ac.uk/Documents/Reports/JISC%20Collections%20OA%20Fees%20Final%20Report.pdf>.
- [21] AAAS announces open-access journal [EB/OL]. [2014-10-17]. <http://blogs.nature.com/news/2014/02/aaas-announces-open-accessjournal.html>.

- [22] LAAKSO M, BJÖRK B C. Anatomy of open access publishing: a study of longitudinal development and internal structure [EB/OL]. [2014-10-17]. <http://www.biomedcentral.com/1741-7015/10/124/>.
- [23] Growth of Fully OA Journals Using a CC-BY License [EB/OL]. [2014-10-17]. <http://oaspa.org/growth-of-fully-oa-journals-using-acc-by-license/>.
- [24] 科学出版将迎来“新的一页”(二):免费开放获取成为科学期刊出版新潮流[EB/OL]. [2014-10-17]. http://digitalpaper.stdaily.com/http_www.kjrb.com/kjrb/html/2014-05/03/content_259689.htm?div=-1.

作者简介

王星, 男, 1977年生, 中国科学技术信息研究所工程师, 研究方向: 数据挖掘与知识服务。

苏静, 女, 1988年生, 武汉大学信息管理学院博士生, 研究方向: 知识组织与语义出版, E-mail: owensujing@163.com。

杨代庆, 男, 1975年生, 中国科学技术信息研究所高级工程师, 研究方向: 信息资源建设。

Analysis of Foreign Open Science and Technology Information Resources Development

WANG Xing¹, SU Jing², YANG DaiQing¹

(1. Institute of Scientific and Technical Information of China, Beijing 100038, China;

2. School of Information Management, Wuhan University, Wuhan 430072, China)

Abstract: In an open information environment, open science and technology information resource has become a key part of academic exchange ecosystem, and is an essential element of prompting modern scientific production. This paper selects 5 foreign science and technology information resources (open journals, open books, open conference resources, open dissertations, open scientific data) as research samples and analyses the status quo as well as construction features. Based on these conditions, it explores some practical issues concerning open resources unified management at the national level, the system for long-term preservation of open source, and open content integration from multiple open platforms.

Keywords: Open access; Science and technology information resources; Information resource development; OA resource assessment

(收稿日期: 2014-10-24)

NSTL全国开通电子资源

NSTL以“国家许可”的方式引进数字资源, 最大程度地实现了规模效益, 节约国家在文献资源经费的整体投入, 重点弥补国内文献资源的历史性缺失。

自2002年引进第一个全国开通的现刊数据库MANEY开始, 截至目前, NSTL以“国家许可”的方式共引进了40个现刊数据库及3个事实性数据库, 2014年开通期刊数量达到655种。重点保障国际知名中小学协会的期刊, 如美国植物生物学家学会、北美放射学会、美国营养学会、美国遗传学学会、美国植物学会、美国冷泉港实验室、美国医师协会、美国胸科医师学会和美国神经放射学会等期刊均极大满足了国内科研人员的需求。

自2008年至今共引进回溯期刊数据库16个, 包括Springer、OUP、IOP、RSC、LWW、BMJ和OUP等国际知名出版社、学协会, 涵盖期刊品种数量达到1500余种, 约600万篇期刊全文回溯内容。

中国大陆非盈利学术机构通过向NSTL提交开通申请(申请流程见<http://www.nstl.gov.cn/全文文献栏目>), 均可以免费使用以上资源。机构用户需通过出版社平台访问现刊数据库内容, NSTL平台只提供链接及跨库检索服务。机构用户均可通过NSTL自建的回溯平台(<http://archive.nstl.gov.cn/Archives/>)访问所有已购买的回溯资源。