

开放获取期刊可达性和稳定性的调查与研究 ——以 DOAJ 收录期刊为例

■ 罗 新

收稿日期:2017-09-03

修回日期:2017-12-07

华南理工大学工商管理学院,广东省广州市天河区五山路 381 号 510641

摘要 【目的】调查开放获取期刊目录(DOAJ)收录的所有开放获取(OA)期刊的可达性和稳定性,以期为OA期刊长期保存和访问提供技术支持和参考。【方法】采用网络调研法,结合Excel宏和人工测试,对2017年5月8日下载的DOAJ元数据中全部期刊进行可达性和稳定性的调查。该调查为期3个月,分3次进行。【结果】在3次调查中,能够成功访问至少1次的期刊共9104种,占样本期刊总数的96.4%。虽然第1次和第2次调查的可达率接近,但是从单个期刊角度分析,在第1次调查中可达的期刊,在第2次调查中未必可达。【结论】DOAJ收录的OA期刊可达性较高,但其稳定性还有待提高。经济发达的国家或地区的OA期刊数量较多、可达性较高。虽然各学科OA期刊数量差异较大,但是各学科期刊可达性差别不大。

关键词 开放获取;开放获取期刊目录;可达性;开放获取期刊;网络引文

DOI:10.11946/cjstp.201709030749

开放获取(Open Access, OA)是在网络环境下发展起来的新型学术交流理念和交流机制^[1-2]。随着网络的普及,OA期刊促进了信息传播,推动了学术繁荣,已经成为重要的学术资源。OA期刊的快速发展,离不开世界各国政府、国际组织、科研机构、出版商以及图书馆的共同支持^[3-4]。2014年,李克强总理在全球研究理事会北京大会上的致辞中强调“中国奉行互利共赢的开放战略”“支持建立公共财政资助的科学知识开放获取机制,促进中国和世界科学事业共同发展”。以此为契机,我国在OA方面也取得了值得肯定的成果^[5]。

OA期刊的基本形式为电子期刊,以网络获取为主,不像纸质文献具有有形的实体可以保存。因此,需要特定的方案来保证OA期刊长期、稳定地被访问。早在20世纪90年代,国外已经正式开展了关于数字文献资源长期保存、长期访问的研究活动。该领域的研究已经走过了基础理论探讨、个体实验、实践3个过程^[6]。目前,国外研究人员还在为完善数字资源的长期保存机制而继续努力^[7]。现阶段相对成熟的数字资源长期保存、访问的方案为LOCKSS、CLOCKSS和Portico^[8-9]。LOCKSS(Lots of Copies Keep Stuff Safe)于1999年开始运行,采用联盟制,会员分为电子期刊出版商和图书馆2类。其

中,电子期刊出版商既包括STM(Science, Technology and Medicine)大型出版社,也包括大量开放获取期刊目录(DOAJ)小型出版商。CLOCKSS(Controlled LOCKSS)是LOCKSS的改进版,该方案优选出小部分图书馆和出版商来完成电子期刊保存任务,而不是像LOCKSS一样由全部成员共同承担保存任务。Portico开始于2002年,采用第三方资源保存机构来解决电子期刊的保存问题。2008年11月,数字图书馆高层论坛首届年会——《国家数字文献资源长期保存服务体系建设》在北京召开,会议正式启动了我国数字文献资源的长期保存体系的建设。但国内对于数字文献长期保存、长期访问的研究还相对滞后^[10]。

虽然国际上已经提出了各种各样关于电子资源长期保存、长期访问的方案,但是针对OA期刊的方案非常少。DOAJ(<https://doaj.org/>)是由瑞典隆德大学(Lund University)图书馆于2003年主办的一个组织、检索OA期刊的工具。DOAJ作为目前国际上最综合的一个OA期刊索引,只有8%的DOAJ期刊被LOCKSS或者CLOCKSS保存,在Portico中保存的更少,所占比例仅为5%^[11]。如何保障OA期刊能够长期、可靠、稳定地被现代和将来的人类使用,如何促进人类智力成果长期保存和长效利用,日益

基金项目:广州市科学技术协会项目“信息素养科普教育活动”(K2017020201002)。

作者简介:罗新(ORCID:0000-0002-6303-801X),硕士,图书与管理资料检索中心主管,E-mail:luoxin@scut.edu.cn。

成为国际关注的焦点。

各国学者开展了多方面关于文献可达性的研究。较多学者分别研究了医学^[12-14]、图书馆与信息科学^[15-17]、人类与社会科学^[18]、生物信息学^[19]等学科期刊参考文献中网络引文的可达性以及统一资源定位符(URL)衰减率。但上述研究都局限于各自的学科范围,并且网络引文也不特指OA期刊。仅有少数学者开展了包含OA期刊可达性的研究。Agashe等^[20]分析了DOAJ收录的48种商业管理类期刊的国家、语言、主题词表、可达性等特征。Graziotin等^[21]提出OA期刊分析框架,该框架包含18个核心属性,分别为书目信息、活动度量、经济学、可访问性和掠夺性等,并使用该框架系统地分析了30种软件工程和信息系统学科的OA期刊。高丽等^[22]在DOAJ中选取了4种具有国际影响力的OA期刊,以2001—2010年发表的电子期刊论文为数据样本,按照不同学科深入分析了网络引文的引用情况,通过期刊文献半衰期测度,比较研究了OA期刊的利用率。谭贝加^[23]以DOAJ收录的我国医学期刊作为研究对象,分析探讨了OA期刊在认可度、出版时效、网站建设、运行资金、论文质量方面存在的各种问题。然而,这些研究也都是针对特定学科,不是以OA期刊可达性为研究主题。本文将从源头出发,不局限于某门学科、某个国家或者地区、某一特定主题或者文献,以DOAJ元数据包含的所有OA期刊为对象,调查OA期刊的可达性和持续性。

1 调查方法和数据来源

通过3次网络调研来调查DOAJ收录的全部OA期刊的可达性和稳定性。调查对象为DOAJ元数据,该元数据为采用UTF-8编码的CSV格式数据,通过DOAJ官方网站下载。2017年5月8日下载的元数据包含来自118个国家或地区的9443种期刊,涵盖20个学科,共收录论文2497533篇。前2次调研使用Excel宏自动完成,第3次调研利用人工测试。3次调研间隔1个月,分别开始于2017年5月15日,2017年6月15日和2017年7月15日。使用VBA编写Excel宏,运行Excel宏可对DOAJ元数据中每种期刊的“Journal URL”栏进行批量访问。根据RFG2616标准^[24],采用HTTP请求方法中的HEAD请求,HEAD请求可以检查超链接的有效性、可用性和最近的修改。在有限速度和带宽下,

HEAD请求能够更快、更准确地提供调研所需要的信息。HEAD请求可以返回HTTP状态码,通过对返回状态码的统计,分析各OA期刊的访问情况。为了确保调研的准确性,前2次重复调查,以免暂时不可达的OA期刊影响实验结果。整理前2次调研结果,得到暂时不可达OA期刊集合,该集合是第3次调研即人工测试的对象。人工测试可以纠正自动访问引起的结果偏差,比如某些OA期刊出版商使用的软件或者网站服务器可能拒绝Excel宏的HEAD请求。最后的调研结果显示,拒绝Excel宏HEAD请求的情况并不少见,人工测试非常有必要。

2 结果与分析

2.1 研究结果

前2次的调研结果统计见表1。如表1所示,2次自动调研的统计结果非常相似。第1次调研的HTTP请求成功率为55.80%,第2次为56.49%。另外,分别有34.54%和35.19%的“Journal URL”能够在重定向后成功访问。总体而言,分别高达90.34%和91.68%的“Journal URL”能够以某种方式可达;9.66%(912种)和8.32%(786种)的“Journal URL”返回错误信息,无法成功访问。

表1 调研1和调研2的结果统计

返回结果	调研1		调研2	
	期刊数量/种	占比/%	期刊数量/种	占比/%
HTTP 2:成功	5269	55.80	5334	56.49
HTTP 3:重定向	3262	34.54	3323	35.19
HTTP 4:请求错误	136	1.44	131	1.39
HTTP 5:服务器错误	83	0.88	104	1.10
其他错误	693	7.34	551	5.83
访问成功小计	8531	90.34	8657	91.68
访问不成功小计	912	9.66	786	8.32
总计	9443	100.00	9443	100.00

3次调研错误代码对比统计分析见表2。如表2所示,错误代码“操作超时”是“Journal URL”不可达的最主要原因,分别占错误总量的50.11%和46.82%。“操作超时”可能由各种各样的原因引起,可能是网站服务器出现故障,也可能是网络断开、使用的软件不兼容,或者程序出错。“操作超时”有可能是暂时的,也可能是永久的。2次结果中,排名第2的原因都是“无法解析服务器的名称或地址”。排名第3~5的访问错误原因数量只有细微的差别。错误代码“403-Forbidden”(服务器理解请求客户端的请求,但是拒绝执行此请求)和“500-Internal Server Error”(服务器内部错误,无法完成请求)表

明,运行 Excel 宏自动访问 OA 期刊的方式可能被期刊服务器限制。这类期刊可能可被成功访问,但是

不允许调研中使用的 Excel 宏访问,第 3 次人工网络调研可以验证该问题,使实验结果更加准确。

表 2 3 次调研错误代码对比统计

错误类型	OA 期刊数量/种					
	N_1	N_{21}	N_{31}	N_2	N_{12}	N_{32}
HTTP 400-Bad Request	1	0	0	1	0	0
HTTP 403-Forbidden	65	13	39	57	2	41
HTTP 404-Not Found	66	5	31	68	10	31
HTTP 406-Not Acceptable	3	0	3	4	1	3
HTTP 416-Requested Range Not Satisfiable	1	0	1	1	0	1
HTTP 500-Internal Server Error	68	15	49	91	31	55
HTTP 502-Bad Gateway	3	1	0	1	0	0
HTTP 503-Service Temporarily Unavailable	12	2	10	12	1	11
URL 无效	4	1	3	3	0	3
操作超时	457	280	9	368	194	5
服务器返回的信息无效或不可识别	36	10	20	32	9	18
无法解析服务器的名称或地址	157	58	3	109	14	1
无法与服务器建立连接	13	11	1	8	5	0
与服务器的连接意外终止	23	5	1	28	8	1
证书中的日期无效或已经过期	1	0	1	1	0	1
证书中的主名称无效或不相符	2	0	1	2	0	1
总计	912	401	172	786	275	172

注: N_1 表示在调研 1 中出现相应错误类型的 OA 期刊数量; N_{21} 表示调研 1 中访问错误的期刊在调研 2 中能够成功访问的 OA 期刊数量; N_{31} 表示调研 1 中访问错误的期刊(排除调研 2 中能够成功访问的期刊)在调研 3 中能够成功访问的 OA 期刊数量; N_2 表示在调研 2 中出现相应错误类型的 OA 期刊数量; N_{12} 表示调研 2 中访问错误的期刊在调研 1 中能够成功访问的 OA 期刊数量; N_{32} 表示调研 2 中访问错误的期刊(排除调研 1 中能够成功访问的期刊)在调研 3 中能够成功访问的 OA 期刊数量。

前 2 次自动调研结果显示,至少能够成功访问 1 次的 OA 期刊共 8932 种,占比 94.59%,即 511 种期刊 2 次均无法成功访问,占比 5.41%。整理这 511 种期刊组成暂时不可达期刊集合,针对该集合进行第 3 次人工测试。为了避免浏览器原因引起的不可访问,在人工测试过程中,交叉使用了 360 安全浏览器 8.1、QQ 浏览器 9.6、Windows Internet Explorer 9.0 3 个浏览器。第 3 次调研成功访问暂时不可达期刊集合中的 172 种期刊,大部分为自动访问中返回“HTTP 403 - Forbidden”“HTTP 404 - Not Found”“HTTP 500-Internal Server Error”“服务器返回的信息无效或不可识别”的期刊,占比 80.81%。

如表 2 所示,部分期刊的访问稳定性较差。在调研 1 中不可达的期刊共 912 种,在调研 2 中能够访问其中的 401 种,在调研 3 中能够访问剩余部分的 172 种。3 次调研后仅有 339 种期刊无法访问,占比 3.59%。虽然调研 1 和调研 2 可达期刊的比例接近,分别为 90.34%和 91.68%,但是可达期刊种类不稳定。例如,调研 1 中因“操作超时”无法访问的期刊共 457 种,在调研 2 中能够访问其中的 280 种,但是调研 2 又新增了 191 种因“操作超时”暂不可达期刊;调研 1 中因“500-Internal Server Error”无法访问的 68 种期刊,在调研 2 中能成功访问其中的 15 种,然而调研 2 又新增了 38 种因同样原因暂不

可达期刊。表 2 也证实了开展本调研的必要性,假如 OA 期刊像传统电子期刊一样,建立长效访问保障机制,则不用担心其潜在的损失风险,不用担心会突然无法访问。

综合 3 次调研结果,共计 9104 种期刊能够成功访问至少 1 次,占比 96.41%,339 种期刊在整个调研过程中始终无法成功访问,占比 3.59%。3 次调研的结果说明,OA 期刊可达性较好,总体可达率达到 96.41%。但是通过交叉分析发现,期刊访问的稳定性还有待提高。

2.2 结果分析

3 次调研都不可达的 OA 期刊共 339 种,分别由 269 个出版商发行,这些出版商来自 59 个国家或地区。总体而言,数量多、访问成功率高的国家或地区大多经济发达。欧洲经济发达,参与 OA 期刊发行也最多,欧洲国民生产总值(GDP)排名前 17 的国家共发行了 OA 期刊 4047 种,占全球总量的 42.86%,而且可达率较高。英国发行了 834 种期刊,发行期刊数量欧洲排名第一,全球排名第二,其可达率超过 99%。瑞士、荷兰、德国也发行了较多 OA 期刊,其可达率都超过 99%。北美洲经济发达,OA 期刊数量也多,美国和加拿大分别发行了 OA 期刊 558 种和 125 种,并且可达率较高。南美洲经济相对较弱,虽然 OA 期刊数量较多,但是可达率排名

相对靠后。巴西发行了 978 种 OA 期刊,位居全球第一,占 DOAJ 元数据中期刊量的 10.36%,但是其可达率只有 91.51%,由于其基数大,调研发现高达 83 种期刊无法访问。墨西哥 OA 期刊数量不多,仅 94 种,其中 25 种期刊无法访问,可达率只有 73.40%。大洋洲 OA 期刊参与度不高。澳大利亚和新西兰分别发行 OA 期刊 84 种和 65 种,澳大利亚 3 种 OA 期刊无法访问。亚洲的印度尼西亚和印度的 OA 期刊数量最多,分别为 582 种和 255 种,但是可达率比同量级的欧洲国家低。中国香港、中国台湾、新加坡、韩国有 OA 期刊加入 DOAJ,其可达率较高,韩国仅有 1 种期刊无法访问。中国大陆地区有 72 种 OA 期刊加入了 DOAJ,全部能够成功访问。超过一半的中东国家没有 OA 期刊加入 DOAJ。中东地区最稳定的伊朗,发行了 301 种 OA 期刊;较稳定的巴基斯坦发行了 49 种 OA 期刊,其可达率处于中游。其他中东国家 OA 期刊数量都很少。非洲经济最发达的经济体——南非和埃及,分别发行了 OA 期刊 62 种和 594 种,全部能够成功访问。

各学科 OA 期刊可达率见表 3。由表 3 可知,OA 期刊总数量最多的为医学类,高达 2239 种,可达率位列第 2。可达率最高的学科为历史辅助学科,可达率为 98.21%。339 种不可达期刊的学科分布非常广泛,涉及到 20 个主学科,173 个子学科,不可达期刊的学科分布较为均匀,没有学科能够全部访问。

表 3 各学科 OA 期刊可达率

学科	期刊数量/种	不可达期刊数量/种	可达率/%
历史辅助学科	56	1	98.21
医学	2239	43	98.08
自然科学	1320	34	97.42
法学	224	6	97.32
地理学·人类学·娱乐	462	15	96.75
农学	422	15	96.45
技术	1011	36	96.44
语言文学	604	22	96.36
美术	204	9	95.59
社会科学	929	44	95.26
教育	700	36	94.86
哲学·心理学·宗教	419	22	94.75
政治学	192	11	94.27
历史(一般)与欧洲历史	209	13	93.78
总论	225	14	93.78
音乐	31	2	93.55
海军科学	14	1	92.86
美洲历史学	54	4	92.59
参考文献·图书馆学·信息资源	102	9	91.18
军事学	12	2	83.33
空值	14	0	
总计	9443	339	

注:空值表示期刊没有标明其所属学科。

3 结语

近几年,我国 OA 期刊虽然取得一定发展,许多刊物相继加入 OA 行列,但是由于起步较晚,与发达国家相比仍然存在不小差距。DOAJ 是世界知名的 OA 期刊目录,被其收录的期刊均为通过严格审核的 OA 期刊。以 DOAJ 元数据为研究对象,开展网络调研,调查了 DOAJ 收录的 OA 期刊的可达性和持续性。调查发现 OA 期刊可达性较高,达到 96.41%,但是访问持续性还有待提高。此外,经济发达的国家或地区,例如北美洲和欧洲,OA 期刊数量较多,可达性也较高。社会不稳定地区,例如部分中东国家,OA 期刊数量较少,可达性也较低。中国大陆地区的 OA 期刊越来越多,目前已有 72 种,并且都能够成功访问;中国大陆、中国香港和中国台湾的 OA 期刊数量总计 136 种,全球排名 20 位,中国台湾仅有 2 种期刊无法访问。各学科 OA 期刊数量差异很大,但是各学科期刊可达率相差不大。因此,提高 OA 期刊可达性和持续性对于提高人类智力成果的利用效率十分重要,值得进一步研究。

参考文献

- [1] 吴文光. 基于 OA 期刊的高校图书馆资源建设与学科服务探索[J]. 大学图书馆学报, 2014, 32(3): 66-68.
- [2] Fecher B, Friesike S, Wagner G G. A nod to public open access infrastructures[J]. *Science*, 2017, 356(6344): 1242.
- [3] 郭进京, 彭乃珠, 张梦霞, 等. 2014 年国际开放获取实践进展[J]. 图书馆工作, 2015, 59(9): 119-126.
- [4] 许洁, 王嘉昀. Open Access 2020 战略背景下的开放获取期刊出版现状——第十二届欧洲学术出版会议综述[J]. 中国科技期刊研究, 2017, 28(7): 593-598.
- [5] 贺郝钰, 侯春梅, 迟秀丽, 等. DOAJ 的运作模式及对中国 OA 期刊的启示[J]. 中国科技期刊研究, 2016, 27(8): 874-879.
- [6] Gust P. Using a digital preservation system to ensure access to online digital content[J]. *Library Hi Tech News*, 2012, 29(9): 5-9.
- [7] Rosenthal D S H. The medium-term prospects for long-term storage systems[J]. *Library Hi Tech*, 2017, 35(1): 11-31.
- [8] Mering M. Preserving electronic scholarship for the future: An overview of LOCKSS, CLOCKSS, Portico, CHORUS, and the Keepers Registry[J]. *Serials Review*, 2015, 41(4): 260-265.
- [9] 孙超, 吴振新. 国外数字资源长期保存成熟度模型及其分析与评价[J]. 图书馆工作, 2017, 61(1): 32-39.
- [10] 吴振新. 数字资源长期保存可信认证研究发展综述[J]. 中国图书馆学报, 2015, 41(3): 114-126.
- [11] Seadle M. Archiving in the networked world: Open access journals[J]. *Library Hi Tech*, 2011, 29(2): 394-404.
- [12] Dellavalle R P, Hester E J, Heilig L F, et al. Going, going, gone: Lost Internet references [J]. *Science*, 2003, 302(5646):

- 787-788.
- [13] Rochon P A, Wu W, Gurwitz J H, *et al.* Prospective evaluation of the accessibility of Internet references in leading general medical journals[J]. *Scientometrics*, 2015, 102(2):1375-1384.
- [14] Nagaraja A, Joseph S A, Polen H H, *et al.* Disappearing act: Persistence and attrition of uniform resource locators (URLs) in an open access medical journal [J]. *Program*, 2011, 45(1):98-106.
- [15] Sadat-Moosavi A, Isfandyari - Moghaddam A, Tajeddini O. Accessibility of online resources cited in scholarly LIS journals: A study of Emerald ISI-ranked journals [J]. *Aslib Proceedings*, 2012, 64(2):178-192.
- [16] Gul S, Mahajan I, Ali A. The growth and decay of URLs citation: A case of an online Library & Information Science journal [J]. *Malaysian Journal of Library & Information Science*, 2017, 19(3):27-39.
- [17] 张丽敏. 基于 CSSCI 网络引文老化实证研究——以图书馆学期刊为例[J]. *情报科学*, 2014, 32(10):114-117.
- [18] Yang S L, Qiu J P, Xiong Z Y. An empirical study on the utilization of web academic resources in humanities and social sciences based on web citations [J]. *Scientometrics*, 2010, 84(1):1-19.
- [19] Wren J D, Georgescu C, Giles C B, *et al.* Use it or lose it: Citations predict the continued online availability of published bioinformatics resources [J]. *Nucleic Acids Research*, 2017, 45(7):3627-3633.
- [20] Agashe A T, Lihitkar S R, Lihitkar R S. Free online journals on business and management on Directory of Open Access Journals (DOAJ) [J]. *SRELS Journal of Information Management*, 2010, 47(1):41-60.
- [21] Graziotin D, Wang X F, Abrahamsson P. A framework for systematic analysis of open access journals and its application in software engineering and information systems [J]. *Scientometrics*, 2014, 101(3):1627-1656.
- [22] 高丽, 王宏光. 基于网络引文的开放存取期刊利用效率研究 [J]. *图书馆理论与实践*, 2014(2):30-33.
- [23] 谭贝加. 中国医学期刊开放获取的现状与分析——以 DOAJ 收录期刊为例 [J]. *中国科技期刊研究*, 2014, 25(9):1113-1116.
- [24] World Wide Web Consortium Network Working Group. Hypertext Transfer Protocol--HTTP/1.1(RFC 2616) [EB/OL]. [2017-05-15]. <http://www.w3.org/Protocols/rfc2616/rfc2616.html>.

Accessibility and stability of open access journals: Taking DOAJ journals as examples

LUO Xin

School of Business Administration, South China University of Technology, 381 Wushan Road, Tianhe District, Guangzhou 510641, China

Abstract: [Purposes] This paper aims to investigate the accessibility and stability of all open access (OA) journals indexed in directory of open access journals (DOAJ), and to provide the basis and references for the research about long-term preservation and access of OA journals. [Methods] We combined Excel macro and manual testing to investigate accessibility and stability of OA journals in the DOAJ metadata downloaded on May 8, 2017 by the network survey method. The survey was conducted 3 times for 3 months. [Findings] Among these 3 surveys, 9104 journals could be accessed at least 1 time, taking up to 96.4%. Although the accessibility rates of the first and second surveys are similar, it can be concluded from the perspective of a single journal that journals accessible in the first survey may not be reachable in the second survey. [Conclusions] The accessibility of the OA journals indexed in DOAJ is high, but the sustainability remains to be improved. In developed countries or regions, the number of OA journals is large, and the accessibility is high. There exists great differences in the number of OA journals in various disciplines, but little difference in accessibility.

Keywords: Open access; Directory of open access journal; Accessibility; Open access journal; Web citation

(本文责编:刘晶晶)