

资源环境领域开放获取仓储目录的分析研究^{*}

杨丽娜, 马建玲, 李慧佳

(中国科学院兰州文献情报中心, 兰州 730000)

摘要: 本文对开放获取仓储目录中登记的195个资源环境学科领域的开放获取仓储从地区分布、语言分布、开放仓储类型、开源软件类型、资源数量、资源类型、互操作性和更新情况等展开系统的分析研究, 讨论该领域开放获取仓储的基本情况和资源特征, 揭示该领域开放获取仓储登记的开放学术资源现状, 最后提出面向该领域开放学术资源建设的对策和建议。

关键词: 开放获取仓储; 开放资源; 资源环境; 资源建设

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2017.09.004

1 引言

开放获取(Open Access, OA)运动的发展,使得越来越多的学术资源可以免费获取。随着开放资源类型的多样化和数量的快速增长,开放资源的学术价值也逐渐提高,已经成为图书馆内除商业资源采购以外的重要馆藏资源^[1]。开放资源获取的主要途径有“开放获取期刊”和“自我存档”,其中“自我存档”即开放获取仓储(Open Access Repositories, OAR),是对数字信息资源的免费存储文档库,其不仅内容免费,而且对使用者没有任何限制,在过去十多年,大部分学研机构、研究中心、政府机构、数据中心或图书馆建立了OAR。OAR收集大量不同类型开放学术信息资源,包括期刊论文、学位论文、图书、会议论文等^[2]。学科知识库专门收集特定学科主题的研究资源,较具有代表性的OAR有高能物理领域的arXiv.org、医学领域的PubMed Central、地球科学领域的Earth-Prints Repository等^[3]。学者们通过对OAR的相关研究,不同程度地揭示了科研人员进行自我存档的态度和行为^[4]、机构知识库引起的图书馆员角色转变^[5]、机构知识库的基础理论以及机构知识库的建设策略^[6]等。

开放获取仓储目录(Directory of Open Access Repositories, OpenDOAR)由英国诺丁汉大学和瑞典伦德大学图书馆于2005年12月联合建立,是全球范围内首家对OAR进行搜集、描述和组织的电子目录,登记数量由建立初期的128个^[7],发展为3 408个(截至2017年8月),该目录具有一定的重要性和权威性。Pinfield等基于2005—2012年的OpenDOAR数据,揭示了世界范围内OAR的快速增长,并概括了OAR的总体特征:以机构知识库为主,多数是跨学科领域,网站登记资源多以英语为主,存在少量OAR登记大量资源和大量OAR登记少量资源现象^[7]。此外,还有学者研究OAR的数量增长、发展趋势及关键特征,分析OAR在促进学术交流中的潜在作用,探讨了机构知识库建设和发展中存在的问题和挑战。一部分学者以地域为基础集中分析OAR的特征,如加拿大^[8]、印度^[9]和中国^[10]等;还有一部分学者重点从学科角度集中分析OAR,如从医学领域^[11]、计算机科学领域^[12]等。

目前还没有学者对资源环境领域OAR展开相关研究,本文将面向资源环境领域开放学术资源的组织利用,从基本情况和资源特征两方面对在OpenDOAR中登记的OAR进行分析。

^{*} 本研究得到中国科学院西北生态环境资源研究院青年人才成长基金项目“资源环境领域开放学术资源的语义集成框架研究”(编号:Y6AJ083013)资助。

2 资源环境领域OAR的基本情况

资源环境领域学科范围广,有多学科交叉性质,涉及生态学、环境科学、地球化学与地球物理学、水资源学等^[13]。在资源环境学科信息门户研究与建设中学科范围以资源与环境科学、地球科学为核心^[14],由此,本文选取OpenDOAR中与资源环境领域密切相关的地球和行星科学(共88个OAR)以及生态和环境学科(共150个OAR)分类下的OAR,去重后得到195个资源环境领域相关的OAR作为分析样本,样本的元数据信息获取时间为2017年5月1—31日。本文从国家分布、类型分布、语言分布以及使用软件四个方面揭示该领域OAR的基本情况。

2.1 国家分布

经统计,OpenDOAR资源环境领域195个OAR分布于55个国家,其中美国、英国、法国等17个发达国家的OAR数量共99个,占总量的一半略强,因此,从总体来看,发达国家和发展中国家OAR的数量相对均衡。Pinfield等研究也表明,OAR发展初期,发达国家数量增长较快,而2010年以后,东亚、南美洲、东欧等地,尤其是巴西、波兰等国的OAR数量开始呈现较快的增长速度^[7]。

2.2 类型分布

本文将资源环境领域OAR的类型分为机构知识库、学科知识库、集成仓储和政府仓储四种,其中机构知识库收集、存放由某个机构或多个学术机构的研究人员产生的学术成果,学科知识库专门收集某一特定学科或主题的研究资源,政府仓储由政府或政府资助机构管理,集成仓储从其他来源收集资源。此外,Nicholas等还补充了OAR的一种类型,即包含特定类型文件的格式存储库,如学位论文、电子期刊、数据集或学习对象^[15]。总体来看,以机构知识库(共151个)和学科知识库(共32个)为主,占比93.8%;集成仓储和政府仓储各有6个,这两种占该领域所有OAR数量的6.2%,这与OAR总体特征^[7]相符。调查表明,全球出版物产出较多的机构,大部分都建立了机构知识库,用以收集、保存和传播自身机构的学术产出^[16]。

在机构知识库方面,国外资源环境领域具有代表

性的机构知识库有佐治亚大学机构知识库^[17]、匹兹堡大学机构知识库^[18]、爱尔兰海洋研究所机构知识库^[19]、海洋生物实验室和伍兹霍尔海洋学研究所机构知识库^[20]等。国内该领域建设机构知识库的相关机构主要集中在中国科学院,包括中国科学院所属的成都生物研究所、地球环境研究所、地理科学与资源研究所、生态环境研究中心、南海海洋研究所、新疆生态与地理研究所、烟台海岸带研究所等;此外,北京大学机构知识库也存储了资源环境领域开放学术资源。在学科知识库方面,Dryad^[21]、CEDA Repository^[22]、PANGAEA^[23]等以存储科学数据为主,其中PANGAEA专门保存、发布和分发来自地球研究系统的地理参考数据;此外,eERL^[24]、OMA^[25]及Earth-prints Repository^[26]等学科知识库都登记和存储了资源环境领域相关的开放学术资源。

2.3 语言分布

资源环境领域的OAR存储了24种语言的开放资源,以多种语言建立OAR,满足不同国家、不同用户的语言需求。总体来看,以英语存储学术资源的OAR有159个,占比81.5%,数量最多,这与Pinfield等的结论相符;其次是西班牙语、法语和中文;登记芬兰语、僧伽罗语等9种语言开放资源的OAR分别有1个。OpenDOAR中,2008年以来登记非英语语种开放资源的OAR比例增加^[7],在资源环境领域,登记除英语外其他语种开放资源的OAR占18.5%。

2.4 使用软件

软件是OAR建设的重要组成部分,资源环境领域OAR建设过程中使用的开源或商业软件共有27种(见表1),其中前11种软件有不少于2个OAR使用,而使用Socionet、panFMP等其他16种软件的OAR分别各有1个,另外还有19个OAR所使用的软件在OpenDOAR元数据中记录为“Not specified”,故未统计在上述软件之列。90个OAR使用最流行的DSpace软件,31个OAR使用EPrints软件,使用这两种开源软件的OAR数量超过一半(占比62.1%)。其他OAR使用的软件如dLibra、OPUS等由于语言限制,仅适用特定国家,dLibra仅波兰的OAR使用,OPUS仅德国的OAR使用。

总体来说,资源环境领域OAR,美国数量最多,发达国家和其他国家的资源数量相对均衡,以机构知识

库和学科知识库为主, 登记的资源有多种语言但主要以英语为主, 大部分仓储使用较为通用的DSpace和EPrints软件。分析表明, 资源环境领域大多数国家的科研机构都积极建设OAR, 在语言和软件上均体现出明显的多样化特征。我国在该领域建设OAR的机构多数集中在中国科学院的相关院所, 且都是机构知识库, 我国没有建设专门的资源环境领域学科知识库, 相比之下资源环境领域建设OAR最多的美国在该领域有6个相关的学科知识库。

表 1 资源环境领域OAR使用软件

序 号	软件名称	数量/个	序 号	软件名称	数量/个
1	DSpace	90	15	eDoc	1
2	EPrints	31	16	DARE	1
3	HAL	9	17	Presto	1
4	Greenstone	6	18	Alma Digital	1
5	dLibra	5	19	invenio	1
6	CONTENTdm	4	20	GX	1
7	Open Repository	4	21	CWIS	1
8	OPUS	4	22	DSpace-CRIS	1
9	Digital Commons	3	23	PHP MySQL	1
10	Drupal	2	24	IMIS	1
11	Fedora	2	25	SciELO	1
12	Socionet	1	26	InfoLib-DBR	1
13	panFMP	1	27	e-repository	1
14	Inmagic DBTextWorks	1	28	-	19

3 资源环境领域OAR的资源特征

本文统计分析的195个资源环境领域OAR共登记约490万条开放资源, 包括期刊论文、学位论文、图书、会议论文等各种类型的学术资源, 分析OAR登记的学术资源的数量、类型、互操作性以及更新情况, 对资源环境领域OAR的开放资源再利用具有实践指导意义。

3.1 资源数量

资源环境领域OAR的资源数量范围分布情况如下: 登记资源数量小于0.1万的OAR有30个, 占比15.4%; 登记资源量在0.1万—0.5万的OAR有70个, 占比35.9%; 登记资源量在0.5万—1万以及1万—5万的OAR均有37个, 各占19.0%; 登记资源数量大于5万的OAR

有17个, 占比8.7%; 另外, 有4个OAR的OpenDOAR元数据信息中资源数量为空白, 因此没有统计在上述资源数量范围之列。对比Loan对OpenDOAR中医学领域OAR的分析^[14], 医学领域登记资源量大于5万的OAR有11个, 占该领域OAR (共254个) 的4.3%, 对比发现, 资源环境领域登记大量资源的OAR相对较多。

3.2 资源类型

资源环境领域OAR登记的资源类型分布如表2所示, 共有11种类型的开放资源, 包括常规开放资源 (期刊论文和会议论文等) 和领域特色资源 (数据集和地图等)。其中, 收集期刊论文的OAR最多 (占比77.4%), 其次是未发表的资源、学位论文、会议论文、专著、多媒体资源、参考资料、专门资源、学习对象资源、专利、数据集。其中专门资源包含地图、公告、灰色文献、新闻、科技报告等。根据统计发现, 资源环境领域大多数OAR都存储了多种类型的开放资源, 仅收录一种资源的OAR相对较少, 这与近年来开放获取范畴逐渐扩大的趋势相同, OAR登记的开放资源类型多样, 相对较全面。登记的资源类型中, 期刊论文、会议论文、学位论文、图书可以较直接反应学术成果的资源相对较多, 登记在OAR中的多媒体资源、专门资源、学习对象资源、数据集等资源比例相对较少。

表 2 资源环境领域OAR的资源类型

排 名	资源类型	数量/个	占比/%
1	期刊论文	151	77.4
2	未发表的资源	96	49.2
3	学位论文	94	48.2
4	会议论文	94	48.2
5	专著	86	44.1
6	多媒体资源	51	26.2
7	参考资料	43	22.1
8	专门资源	38	19.5
9	学习对象资源	37	19.0
10	专利	18	9.2
11	数据集	14	7.2

数据集资源以及专门资源中包含的地图充分体现了资源环境领域的特色。登记数据集资源的OAR有14个, 其中8个是学科知识库、7个是机构知识库, Dryad、PANGAEA、CEDA Repository、Earth-prints Repository^[27]

等重要学科知识库都收集和存储了数据集资源,尤其是亚利桑那大学地球科学系的学科知识库RRUFF Project^[28]只登记数据集这一类资源。

3.3 互操作性

大部分OAR遵循OAI-PMH协议,能够被专业搜索引擎收录,拓宽了资源的检索途径^[29]。OAI-PMH协议支持OAR资源的元数据共享和互操作,增强了系统的互操作性。OpenDOAR中资源环境领域有133个(占比68.2%)OAR提供OAI-PMH接口的链接地址,支持用户通过OAI-PMH协议获取仓储中资源的元数据内容;而31.8%的OAR不提供OAI-PMH接口的链接地址,其登记的开放资源的使用和获取权利在具体每个OAR的服务平台中都有相应的政策和说明。

3.4 更新情况

OAR登记资源的持续更新,能保证系统的持久性,推动仓储的积极增长和发展。OpenDOAR中OAR的元数据中资源量字段的信息2—4周更新一次^[7]。从对2012—2017年发生数据更新的OAR数量来看,2017年有153个资源环境领域OAR登记的学术资源进行了数据更新,2015—2017年有91.8%的OAR对登记数据进行了更新,有4个(占比2.1%)OAR没有提供数据更新年份。通过对数据更新情况的分析,认为资源环境领域OAR对机构新增的开放学术资源进行持续收集和存储,是开放学术资源建设比较可靠的数据源。

总体来看,资源环境领域OAR登记期刊论文、学位论文、数据集等多种类型资源,在数量上已有一定规模并且持续更新,大部分仓储通过提供OAI-PMH接口支持用户对开放资源元数据信息的获取。

4 资源环境领域开放学术资源建设的建议

目前,国内缺乏对资源环境领域开放学术资源的系统化发现、组织和利用实践,本文对OpenDOAR中资源环境领域OAR的分析研究发现,该领域开放学术资源分布在195个OAR中,除期刊论文外,还有许多其他类型资源,开放资源在数量上也具有一定规模,虽然可以免费使用,但因存储分散,使得科研人员没有时间和精力去查找和利用。因此,面对上述问题,结合学科

领域特色,有针对性地进行开放学术资源建设,对不同来源开放学术资源进行系统化集成,能够提升开放学术资源的利用价值,为科研人员提供资源保障和服务。本文结合分析结果,针对资源环境领域开放学术资源的建设提出相关建议。

4.1 整合各种类型资源并突出资源环境领域特色资源

经过长期发展,开放获取的范畴得到延伸,开放学术资源包括期刊论文、图书、科技报告、科学数据、学位论文等各种类型。资源环境领域的开放学术资源建设有必要对OAR中登记的各种类型开放学术资源进行整合,另外,资源环境领域在科研过程的观测和实验中产生大量的科学数据资源,而本文研究的OpenDOAR中有14个OAR在不同程度上登记了科学数据资源,因而在资源建设中应突出开放科学数据资源的集成和保障。OAR中登记的开放资源毕竟有限,为保障该领域开放学术资源建设的全面性,可从OAR登记的资源类型出发,按类型调研和补充其他来源的开放资源。

4.2 资源环境领域OAR中开放资源遴选方法

资源环境领域有195个OAR,在该领域开放资源建设工作中如何遴选和收集所有来源的开放资源非常关键。从本文分析的OAR基本情况和资源特征看,可从国家分布、语言分布、OAR类型、资源数量、资源类型、更新频率、互操作性等特征出发,确立遴选OAR的标准和优先顺序。本文对OAR中开放资源遴选的方法提出两种建议:一是优先收集和整合以英语语言存储开放资源、登记资源数量较多、更新频率快、提供OAI-PMH接口的OAR中的开放资源;二是可仅从资源类型出发,优先对OAR中某一类型资源实施全部收集和保障,如优先收集该领域中存储期刊论文的151个OAR中的期刊论文资源,再依次收集和保障其他类型的资源。

参考文献

- [1] 黄金霞,王昉,张建勇.开放资源的组织和再利用[J].图书情报工作,2016,60(17):116-121.
- [2] Directory of Open Access Repositories[EB/OL].[2017-05-31].<http://www.openoar.org>.

- [3] BJÖRK B C. Open access subject repositories:an overview[J].Journal of the Association for Information Science and Technology,2014,65(4): 698-706.
- [4] KIM J.Faculty self-archiving:motivations and barriers[J].Journal of the Association for Information Science and Technology,2010,61(9): 1909-1922.
- [5] WALTERS T O.Reinventing the library how repositories are causing librarians to rethink their professional roles[J].Portal Libraries and the Academy,2007,7(2):213-225.
- [6] 邱均平,祖旋,郭丽琳,等.机构知识库的研究现状及其发展趋势的可视化分析[J].情报理论与实践,2015,38(1):12-17.
- [7] PINFIELD S,SALTER J,BATH P A,et al. Open-access repositories world wide,2005—2012:past growth,current characteristics,and future possibilities[J].Journal of the Association for Information Science and Technology,2014,65(12):2404-2421.
- [8] SHEARER K.The CARL institutional repositories project:a collaborative approach to addressing the challenges of IRs in Canada[J].Library Hi Tech,2006,24(2):165-172.
- [9] SAWANT S.Institutional repositories in India:a preliminary study[J].Library Hi Tech News Incorporating Online and Cd Notes,2011,28(10):6-10.
- [10] 聂银,刘锦宏,张榕洁,等.我国开放获取仓储分布研究[J].科技与出版,2013(4):101-105.
- [11] LOAN F A,SHEIKH S.Analytical study of open access health and medical repositories[J].Electronic Library,2016,34(3):419-434.
- [12] BHAT M H.Interoperability of open access repositories in computer science and IT—an evaluation[J].Library Hi Tech,2010,28(1):107-118.
- [13] 张鸿翔,杨鹏,赵千钧.关于中国科学院资源环境领域人才问题的几点思考[J].中国科学院院刊,2014,19(6):457-461.
- [14] 祝忠明,吴新年,孙成权,等.资源环境学科信息门户的研究与建设[J].图书情报工作,2005,49(7):65-69.
- [15] NICHOLAS D,ROWLANDS I,WATKINSON A,et al.Digital repositories ten years on: what do scientific researchers think of them and how do they use them?[J].Learned Publishing,2012,25(3):195-206.
- [16] 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.
- [17] Athenaeum@UGA[EB/OL].[2017-08-08].<http://athenaeum.libs.uga.edu/>.
- [18] D-Scholarship@Pitt[EB/OL].[2017-08-08].<http://d-scholarship.pitt.edu/>.
- [19] Marine Institute Open Access Repository(OAR)[EB/OL].[2017-08-08].
<http://oar.marine.ie/>.
- [20] Woods Hole Open Access Server(WHOAS)[EB/OL].[2017-08-08].
<https://darchive.mblwhoilibrary.org/>.
- [21] Dryad[EB/OL].[2017-08-08].<http://www.datadryad.org/>.
- [22] CEDA Repository[EB/OL].[2017-08-08].<http://cedadocs.badc.rl.ac.uk/>.
- [23] PANGAEA[EB/OL].[2017-08-08].<http://www.pangaea.de/>.
- [24] Electronic Environmental Resources Library(eERL)[EB/OL].[2017-08-08].
<http://www.eerl.org/index.php>.
- [25] Open Marine Archive(OMA)[EB/OL].[2017-08-08].<http://www.vliz.be/en/open-marine-archive>.
- [26] Earth-prints Repository[EB/OL].[2017-08-08].<http://www.earth-prints.org/>.
- [27] 王丹. OAR发展研究[D].长春:东北师范大学,2011.
- [28] RRUFF Project[EB/OL].[2017-08-01].<http://rruff.info/>.
- [29] 陈美华,刘文云,刘昊,等.国内外机构知识库建设研究[J].情报理论与实践,2015,38(9):55-63.

作者简介

杨丽娜,女,1991年生,硕士,助理馆员,研究方向:开放知识资源建设与组织,E-mail: yangln@llas.ac.cn。
马建玲,女,1969年生,硕士,研究馆员,研究方向:开放知识资源建设与组织,E-mail: majl@lzb.ac.cn。
李慧佳,女,1984年生,硕士,馆员,研究方向:开放知识资源建设与组织,E-mail: lihj@llas.ac.cn。

Analytical Study of Open Access Repositories in Resource and Environment Field

YANG LiNa, MA JianLing, LI HuiJia
(Lanzhou Library Chinese Academy of Sciences, Lanzhou 730000, China)

Abstract: In this paper, the 195 resources and environment open access repositories registered in the Directory of Open Access Repositories were discussed deeply in the aspects of regional distribution, language distribution, repositories type, software, resource quantity, content type, interoperability, and updated situation. The study revealed the basic situation and resource characteristics of open access repositories in the resources and environment field. The study also disclosed the current situation of open resources stored in the open access Resources and Environmental Repositories. Finally, the paper gave suggestions about the construction of open academic resources in this field.

Keywords: Open Access Repository; Open Resources; Resources and Environment; Resource Development

(收稿日期: 2017-08-09)