

●陈传夫 王云娣

机构典藏资源的分布及获取策略研究*

摘 要 机构典藏库(IR)是由学术机构建立的,以网络为依托,以收集、整理、保存、检索、提供利用为目的,以本机构所创建的数字化产品为内容,供机构内外的终端用户利用,几乎没有任何使用限制的知识库。目前,大部分的 IR 分布在科学技术发达的国家。IR 作为学术机构的智力馆藏,资源内容包括正式出版物和非正规渠道得到的灰色文献两部分。IR 资源的获取方式有直接获取法、间接获取法和集成目录获取法。表6。参考文献6。

关键词 机构典藏库 数字产品 开放存取 信息获取

分类号 G250.73

ABSTRACT Institutional repositories are created by academic institutions, based on networks, with the objectives of acquisition, arrangement, preservation, access and utilization, with the digital products created by people within the institutions as their major contents, provided for use by both inside and outside users, and without any use limitation. At present, most IRs are distributed in technologically advanced countries. As intellectual collections of academic institutions, IR contents include both publications and grey documents. The acquisition methods for IR include direct, indirect and integrated-catalog acquisition methods. 6 tabs. 6 refs.

KEY WORDS Institutional repository. Digital product. Open access. Information access.

CLASS NUMBER G250.73

机构典藏库(Institutional Repositories, IR)是由学术机构建立的,以网络为依托,以收集、整理、保存、检索、提供利用为目的,以本机构所创建的各种数字化产品为内容,供机构内外的终端用户利用,几乎没有任何使用限制的知识库^[1-3]。

伴随着 Depace 等一些项目的启动,机构典藏这一概念逐渐被人们接受。它为传统的学术交流提供了新形式,极大丰富了用户可利用的资源,为数字环境下用户的信息交流和知识共享提供了新的平台。了解网上机构典藏资源的成长与分布情况,对加快和扩大网上信息资源开发利用的进度和规模,利用网络共享信息资源,弥补图书情报机构普遍面临的信息资源短缺问题,有非常重要的现实意义。

1 机构典藏库的成长及分布研究

通过对 The Directory of Open Access Repositories-OpenDOAR 网站的检索发现,网上已有相当丰富的机构典藏资源^[4]。到 2006 年 11 月 12 日,OpenDOAR 网站已收录机构典藏库 827 个。

1.1 机构典藏库的成长情况

为详细地了解机构典藏库的成长情况,我们访问了 Registry of Open Access Repositories (ROAR) (<http://roar.eprints.org>) 网站,随着时间的变化,在该网站登记的 IR 的数量不断增长,据“Repositories Registered Over Time”⁵ 的显示我们可以看出,1990 年至 1998 年期间,国际上机构典藏库增长势头迅猛,这与国际社会普遍开展的开放存取运动有密切关系,说明开放存取促进了机构典藏库的发展。

1.2 机构典藏库的数量分布

OpenDOAR 网站收录的 827 个机构典藏库分布在 44 个国家和地区(见表1)。大部分集中在科学技术发达的国家,如美国、德国、英国、澳大利亚等国。尤其是美国,共有 251 个,占总量的 30.35%。其次是德国,共有 110 个,占 13.30%。DOAR 网站收录我国的有 2 个,分别是 Hong Kong University of Science and Technology Institutional Repository (HKUST Repository) 和 Kong Kong University Theses Online (HKUTO)。

* 本文系高等学校博士学科点专项科研基金资助课题成果之一(项目编号 20050486055)。

表1 机构典藏库的数量分布 (个)

国家	数量	国家	数量	国家	数量	国家	数量
美国	251	日本	16	挪威	5	哥斯达黎加	1
德国	110	比利时	13	奥地利	4	克罗地亚	1
英国	83	西班牙	11	匈牙利	4	韩国	1
澳大利亚	50	新西兰	9	爱尔兰	4	纳米比亚	1
荷兰	44	波兰	8	墨西哥	4	巴基斯坦	1
加拿大	29	芬兰	7	葡萄牙	4	秘鲁	1
法国	28	丹麦	6	中国	2	菲律宾	1
瑞典	28	南非	6	哥伦比亚	2	俄罗斯联邦	1
巴西	25	瑞士	6	捷克	2	塞尔维亚	1
意大利	24	委内瑞拉	6	希腊	2	新加坡	1
印度	16	智利	5	以色列	2	斯洛文尼亚	1

1.3 机构典藏库的学科分布

48.49%,居首位。其次是计算机和信息技术的为71

OpenDOAR 网站收录的827个机构典藏库涉及个,占总数的8.59%,详见表2。

29个学科,其中涉及多个学科的有401个,占总数的

表2 机构典藏库的学科分布 (个)

学科	数量	学科	数量	学科	数量
多学科	401	物理学、天文学	45	哲学、宗教	27
计算机、信息技术	71	法律、政治	44	地球科学	23
健康、医学	66	生物学、生物化学	40	语言、文学	23
自然科学	64	数学、统计学	40	心理学	20
工程技术	63	生态学、环境科学	31	电学、电工	18
历史、考古	63	教育学	29	农业、食品、兽医	17
社会科学	63	化学、化工	28	经营、管理	14
商业、经济	53	机械工程、材料	27	建筑学	13
地理学	49	人文科学	27	城市规划	11
图书馆、信息科学	49	表演艺术	27		

1.4 机构典藏库的内容类型分布

为14种类型,具体包括:学位论文、会议文献、图书、

机构典藏库作为学术机构的数字馆藏,其资源内容多媒、后印本、预印本、学习对象、参考工具书、数据等(见表3),有的涉及多种内容类型。

文献两部分。OpenDOAR 对机构典藏库按内容划分

表3 机构典藏库的内容类型分布 (个)

类型	数量	类型	数量	类型	数量	类型	数量
Theses	441	Books	259	References	130	Software	22
Unpublished	400	Multimedias	171	Learning Objects	113	Patents	15
Publications	306	Postprints	146	Preprints	60		
Conferences	293	Special	133	Datasets	57		

1.5 机构典藏库的语种分布

语、法语和西班牙语,多数采用两种及以上语言出版

OpenDOAR 网站收录的827个机构典藏库涉及(见表4)有7个机构典藏库除用英语还用汉语出版,分别为:Connexions Repository (<http://cnx.org/>),

HKUST Repository (<http://repository.ust.hk/>) HKUTO (<http://dspace.anu.edu.au/>), JHSPH Open Course (<http://sunzil.lib.hku.hk/hkuto/>), Ibiblio (<http://www.ibiblio.org/>), Ware (<http://ocw.jhsp.hk.edu/>) 和 QSpace at Queen's University (<http://qspace.library.queensu.ca/>)。

表4 机构典藏库的语种分布

语言	数量	语言	数量	语言	数量	语言	数量
英语	718	日语	17	拉丁语	4	捷克语	1
德语	124	波兰语	8	俄语	4	希伯来语	1
法语	55	汉语	7	北印度语	3	埃纳德语	1
西班牙语	45	芬兰语	6	朝鲜语	3	塞尔维亚语	1
荷兰语	38	加泰罗尼亚语	5	匈牙利语	2	斯洛文尼亚语	1
瑞典语	28	挪威语	5	毛利语	2	乌尔都语	1
意大利语	26	丹麦语	4	南非语	1		
葡萄牙语	19	希腊语	4	克罗地亚语	1		

1.6 机构典藏库的软件使用分布

根据 OpenDOAR 网站对 IR 系统使用软件的统计,我们发现,827 个机构典藏库中,有 236 个没有注明所使用的软件,其余 591 个使用的软件多达 57 种,

以 GNU Eprints 软件进行构建的典藏库数量位居前列,共有 175 个,占总数的 21.16%,其次为 DSpace 进行构建的 IR 为 148 个,位居第二,详见表 5。

表5 机构典藏库的软件使用分布

软件类型	数量	软件类型	数量	软件类型	数量	软件类型	数量
GNU Eprints	175	eDoc	3	DigiTool	1	perseus	1
Dspace	148	Cocoon	2	DLXS	1	PMC	1
Bepress	54	CWIS	2	Dynawen	1	PORT	1
OPUS	48	Fedora	2	ePub	1	SirsiDunix	1
Wildfire	23	i-Tor	2	ETU-Bibliothek	1	STAR Database	1
HTML	20	Miless	2	Fez	1	Sunsite	1
ETD-db	15	MyCoRe	2	HS-DVL	1	UBKA	1
Diva-Portal	14	OAI Cat	2	Hyperwave	1	Ultraseek	1
CDSWare	8	Plone	2	MANITOU	1	Update Software	1
CONTENTdm	8	RedHat	2	Maxwell	1	VITAL	1
HAL	7	ARCHIMEDES	1	Mercury	1	VTs	1
dLibra	6	arXiv	1	MONARCH	1	XoonIps	1
Open Repository	5	D-Scholarship	1	NPS	1	not specified	236
Cybertesis	4	DARE	1	OJS	1		
DOKS	3	Digital Commons	1	OpenCourseWare	1		

注:表1~表5的调查时间为2006年11月12~13日。

2 机构典藏库资源的获取

2.1 直接获取法

在因特网上,若是知道机构典藏库的网址,通过直接写入网址查询的方法,称之为直接获取法。如已知加利福尼亚理工学院电子论文和学位论文典藏库(Caltech Electronic Theses and Dissertations)的网址,通过写入 <http://etd.caltech.edu/ETD-db/ETD->

[browse/browse?first-letter=all](http://etd.caltech.edu/ETD-db/ETD-browse/browse?first-letter=all),就可以直接查到该机构典藏库。再如,香港科技大学机构典藏库的网址为 <http://repository.ust.hk/dspace/>,通过在浏览器的地址栏中键入该网址,就可访问该典藏库的资源。用户可通过浏览(包括:团体和收藏、主题、作者、日期)和检索(快速检索和高级检索)来获取自己所需要资源。

2.2 间接获取法

不知道具体网址时,通过搜索引擎的帮助,找到所需的机构典藏库,称为间接获取法。我们可利用综合通用型搜索引擎,如: Alta Vista, Lycos, Google, Hotbot 等,其主页面上都提供关键词检索,而且这些检索都含有简单检索和复杂检索两种方式。用户只要在搜索框中输入检索词或检索表达式,如“Institutional Repository”, “Research Repository” “University * Repository”, “OpenAccess Repositories”等,搜索引擎就会返回一组指向相关站点的超链接。由于是机器人程序自动搜索,可方便地收集更多的网站并及时更新、发现和删除已不存在的站点,提高查询结果的数量。

还可以利用专门性的开放存取搜索引擎 OAIster (<http://www.oaister.org/>) 和 Citebase (<http://citebase.eprints.org/cgi-bin/search>) 进行检索。例如:由美国密执安大学数字图书馆制作服务部承担的 OAIster 项目,是一个提供电子书、电子期刊、录音、图片及电影等数字化资料“一站式”检索的门户网站,被美国图书馆协会评为 2003 年度最佳免费参考网站。资料源自 700 多家机构,用户可按字顺浏览机构,每个机构下面列出了所收录记录的数量、访问的网址和简单的描述信息。用户还可按关键词、题名、创作者、主题或资源类型进行检索。检索结果含资源描述和该资源的链接。标引对象包括国会图书馆美国记忆计划、各类预印本及电子本文献服务器、电子学位论文,共有记录 970 多万条^[5]。

2.3 集成目录获取法

机构典藏库作为开放存取资源的重要组成部分,大部分支持元数据收割协议 OAI-PMH,通过到开放存取登记处注册,被统一集成。网上已有许多机构典藏库的集成目录网站,通过对这些站点的访问,可以方便地获取相关机构典藏资源,起到事半功倍的效果。

OpenDOAR(网址:<http://www.opendoar.org/>)是开放存取典藏库名录的缩写,全称是 Directory of Open Access Repositories,由英国诺丁汉大学和瑞典 Lund 大学合作承担,DOAR 对各领域的开放存取典藏库进行记录和分类。过去的两年,各种开放存取典藏库如雨后春笋般建立起来并迅速发展,DOAR 的出现,适应了研究人员方便地获取开放存取资源的要求。DOAR 提供综合权威的机构典藏库、学科典藏库和由各种资助机构建立的开放存取典藏库(如 NIH, Wellcome Trust)列表。用户可以通过典藏库的地点、类型、收藏资料等方式检索使用这些典藏库。DOAR 既适合那些希望在专门的典藏库中查找原始研究论文的用户使用,也适合诸如搜索引擎等服务第三方使用。

Registry of Open Access Repositories (ROAR, <http://roar.eprints.org/>)是一个开放存取机构库注册的网站,由英国南安普敦大学的 Tim Brody 编制维护,目前已收录各种类型的机构典藏库 780 多个,是获取机构典藏库资源的重要网站,用户可通过国家、内容类型、所使用软件的类型来浏览所需要的机构典藏库资源,该网站还提供快速检索功能,方便用户检索。

Registered Data Providers (<http://www.openarchives.org/Register/BrowseSites>)是一个收录依 OAI-PMHv2.0 协议构建的机构典藏库的登记网站,到 2006 年 12 月 4 日,已收录 546 个典藏库,按典藏库的名称、URL 排序,用户可通过浏览 IR 的名称,进入相关机构典藏库的网站。

OAI Repository Explorer(<http://re.cs.uct.ac.za/>)由开普敦大学的 Hussein Suleman 建立维护,目前收录机构典藏库 108 个。

通过上述方法的综合运用,笔者对搜索到的机构典藏库网站进行筛选后,列出国内外一些典藏库的名称、网址、记录数量及可获取性的情况(见表 6)

表 6 国内外主要的 IR 网站资源

国家	名称	网址	记录数量及可获取性
美国	DSpace at MIT	http://dspace.mit.edu/	21012 70% 的全文可免费获取
美国	Cornell University DSpace Digital Repository	http://dspace.library.cornell.edu/index.jsp	不详 100% 可免费获取全文
日本	National Institute of Informatics	http://ju.nii.ac.jp/	123969 100% 可免费获取全文

续表

国家	名称	网址	记录数量及可获取性
德国	MeIND;Hauptseite CDSWare	http://www.meind.de/	30981 100% 可免费获取全文
德国	Ludwig-Maximilians-Univer Sität München Dissertationea	http://edoc,Ub.Uni-muenchen.de/	3070 100% 可免费获取全文
荷兰	Wageningen Univetsity and Research Centre	http://library.wur.nl/way/	112838 50% 可免费获取全文
英国	CCLRC ePublication/Archive	http://epubs.cclrc.ac.uk/	23266 30% 可免费获取
英国	DSpace at Cambridge	http://www.dspace.cam.ac.uk/	179692 100% 可免费获取全文
法国	HAL;Hyper Article en Ligne	http://hal.ccsd.cnrs.fr/	34338 100% 可免费获取全文
法国	Revues.org Fédération de revues en ligne en sciences humaines et sociales	http://www.revues.org/	9996 50% 可免费获取全文
巴西	Biblioteca Digital de Tesese e Dissertacoes (BDTD)	http://bdttd.ibict.br/bdttd	30200 70% 可免费获取全文
澳大利亚	The Australian National University	http://dspace.anu.edu.au/	43964 100% 可免费获取全文
瑞典	Lund University Institutional Archive; LU; research GNU EPrints	http://ask.lub.lu.se/	6241 30% 可免费获取全文
加拿大	T-Space at The University of Toronto Libraries	https://tspace.library.utoronto.ca/	9055 100% 可免费获取全文
西班牙	RACO	http://www.raco.cat/	20908 75% 可免费获取全文
中国(香港)	Hong Kong University of Science and Technology Institutional Repository	http://repository.ust.hk/	2290 100% 可免费获取全文
印度	Indian Institute of Science. Bangalore. India	http://eprints.iisc.ernet.in/	2494 70% 可免费获取全文
意大利	European University Institute	http://cadmus.iue.it/dspace/	2193 100% 可免费获取全文
中国(香港)	Hong Kong University Theses Online (HKUTO)	http://sunzil.lib.hku.hk/hkuto/index.jsp	2699 100% 可免费获取全文

注:上述 IR 网站资源采集时间为 2006 年 11 月 12 日。

参考文献

- 1 Clifford A. Lynch. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL, 2003 (226). [2006-11-03]. <http://www.arl.org/nwealter/226/ir.html>
- 2 Ware, Mark. Pathfinder Research on Web-based Repositories. [2006-11-03]. <http://www.palagroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67ca80256e370051e88a?OpenDocument>
- 3 Crow Raym. The Case for Institutional Repositories: A SPARC Position Paper. [2006-11-03]. <http://www.arl.org/SPARC/IR/ir.html>
- 4 OpenDOAR-OpenAccess Repositories. [2006-11-12]. [tp://www.opendoar.org/](http://www.opendoar.org/)
- 5 Generetea Chart (ROAR). [2006-03-31]. <http://roar.eprints.org/index.php?action=generate-chart>
- 6 OAister. [2006-11-14]. <http://oaister.umdl.nmich.edu/o/oaister/>

陈传夫 武汉大学信息管理学院教授。通讯地址:武汉市。邮编 430072。

王云娣 浙江师范大学文献信息研究中心研究馆员。通讯地址:浙江金华浙江师范大学图书馆。邮编 321004。

(来稿时间:2006-10-26)