

日本机构知识库的发展及对我国的启示

王艳霞 叶小梁

【摘要】机构知识库作为开放获取的一种实现模式，在世界范围内尤其是欧美国家发展迅速，而在亚洲只有日本发展最为迅速。文章首先对日本机构知识库发展的总体情况、政策层面的支持及其在线机构知识库进行介绍，然后对日本机构知识库的建设主体、资源类型及语言分布情况进行详细的分析，最后有针对性地对我国机构知识库的发展提出建议。

【关键词】日本 机构知识库 JAIRO

Abstract: As an implementation pattern of open access, institutional repository has developed rapidly in the world, especially in the west. And in Asia, Japan is the fastest growing nation. This paper describes the development, policy support and the JAIRO of Japanese institutional repository. Then the main building, resource type and language distribution of institutional repository in Japan are analyzed. Finally, issues for the development of Chinese institutional repository are put forward.

Key words: Japan institutional repository JAIRO

机构知识库，也称为机构仓储 (Institutional repository, 简称 IR)。关于机构知识库的定义，迄今为止还没有统一的认识。作为开放获取的一种实现模式，Lynch 从大学的角度对其定义如下：大学中的机构知识库是大学为其员工提供的一套服务，用于管理和传播大学的各个部门及其成员创作的数字化学术科研成果^[1]。我国学者柯平则对其内涵进行了扩展，指出机构知识库是利用网络及相关技术，依附于特定机构而建立的数字化学术数据库，它收集、整理并长期保存该机构及其社区成员所产生的学术成果，并将这些资源进行规范、分类、标引后，按照开放标准语相应的互操作协议，允许机构及其社区内外的成员通过互联网来免费地获取使用^[2]。从目的来看，IR 的建设主要是为了改善信息的传播并提高大学及研究机构的社会价值。

1 日本机构知识库建设概况

1.1 总体发展情况

机构知识库在世界范围内尤其是欧美国家发展迅速，截至 2009 年 5 月 30 日，被开放存取知识库名录 (The Directory of Open Access Repositories, 简称 OpenDOAR) 收录的知识库总数已达 1 407 个^[3]。对亚洲来说，大部分国家的机构知识库发展速度相对缓慢，但日本却是一个例外，2006 年 4 月份日本已经有 13 个机构知识库，而在随后的 3 年内数量已经上升至 109 个 (NII, 2009 年 6 月 13 日)。日本的机构知识库的数量目前位列世界第四，而且将近占亚洲总数的 47%^[4]，是亚洲机构知识库数量最多的国家。

1.2 政策层面支持

2003 年，日本文部科学省的科技理事会在一篇报告中强调大学图书馆在学术信息传播方面发挥着重要作用^[5]。随后，隶属文部科学省的国立情报研究所 (National Institute of Informatics, 简称 NII) 就大力推动日本机构知识库的发展。NII 成立于 2000 年 4 月，主要工作是进行信息学综合研究，同时也是与大学共同开发利用信息资源的机构，既有研究与教育职能，又有事业开发任务。它可以提供国家最大的数据库和联合目录。其独特的地位使其能够获得并发布各个大学的数据和数字化期刊，从而为机构知识库建设提供支持^[6]。为扩大日本开放获取机构的覆盖范围并支持促进学术内容传播的研究活动，NII 于 2006 联合 57 所大学构建“下一代学术内容信息基础设施”，截止到 2007 年这些合作者完成了 22 个项目，如连接链接解析器；综合搜索；开发 IR 评估方法；IR 社区开发；内部期刊出版；交替的开源图书馆模块等等^[7]，从而使整个学术团体受益。在资金方面，仅 2006 年这些项目的总预

算就达 3 亿日元 (约 260 万美元)^[8]。

1.3 日本在线机构知识库——JAIRO

日本在线机构知识库 (简称 JAIRO) 能够对日本各个机构知识库中的学术信息 (如期刊论文、学位论文等) 进行搜索。NII 于 2008 年 10 月 22 日试运行 JAIRO, 2009 年 4 月 1 日对其正式投入使用并将其添加到 NII 学术内容门户 GeNii 综合搜索系统当中^[9]。截止到 2009 年 6 月 13 日, JAIRO 中已经有 113 个组织, 总的资源数目已达 673 653 条。用户通过 JAIRO 既可以集成检索, 也可以按照资源类型方便地检索, 或者直接进入目标机构知识库查找所需内容。通过 JAIRO 机构之间可以共享各个知识库的学术内容, 并且这也促进了各个机构合作进行机构知识库的相关研究。此外, NII 根据各个机构的 IR 管理负责人所提交的相关信息, 收集全国 IR 的元数据构建了机构知识库数据库 (IRDB) 并开发了内容分析系统。通过 NII IRDB 内容分析系统既可以了解各个 IR 的具体信息, 也可以获得 JAIRO 内容的相关数据。本文主要是根据该系统来对日本的 IR 进行相关分析。

2 日本机构知识库分析

2.1 日本机构知识库的创建主体

日本大多数机构知识库都是由高校建立的。NII 的目标是促使日本越来越多的高校都建立机构知识库, 所以自 2004 年日本首次实施机构知识库建设以来, 其数量持续大幅增加。根据 JAIRO 的门户网站 (2009 年 5 月), 目前日本有 97 所大学参与建设机构知识库 (其中广岛知识库和福井知识库是若干个大学联合创建的), 在日本 756 所高校中所占的比例是 12.8%。这些大学有不同的背景, 其地位和支持系统也不尽相同, 大致可以分为 3 组: 87 所国立大学, 89 所县立/府立大学和 580 所私立大学^[10]。机构知识库中在这 3 类大学中的分布是非常不均匀的 (见表 1)。

表 1 各类大学 IR 分布情况

大学	IR 数量	数量 (每组)	建有 IR 的大学所占比例 (每组)
国立大学	66	87	75.90%
县立/府立大学	3	89	3.40%
私立大学	28	580	4.80%
总数	97	756	12.80%

(数据来源: 根据 http://irdb.nii.ac.jp/analysis/index_e.php [2009/05] 整理。)

从上表可以看出, 在大学所建的机构知识库中国立大学所占的比例最大。这可能是由于国立大学有良好的学术环境、技术基础设施和科研设施等条件。虽然私立大学现有知识库数量将近占有所有大学知识库总数的 30%, 但是其建有 IR 的大学占整个私立大学的比例却仅有 4.8%, 远远低于国立大学的 75.9%。对于大部分没有建立 IR 的大学, 他们只有进一步提高图书馆的质量和和服务, 才能逐渐建立起自己的知识库^[11]。自 2009 年以来, 日本的一些研究机构 (如日本国立聚变研究所、贸易振兴机构亚洲经济研究所、网格技术研究中心和奈良文化遗产研究所等) 和县立图书馆、档案馆也逐渐开始建立机构知识库^[12]。NII 也正在鼓励各类学院和大学内部的研究机构建立知识库。

2.2 日本机构知识库收录资料的类型

日本 IR 收录的资料类型分为 14 种 (见表 2)。虽然表面上日本的 IRs 与世界各地 IRs 收录的资料类型相同, 但是仔细分析之后我们会发现这些资料在类型分布和数量方面是有差异的, 而这也反映出了一个国家的学术文化。比如美国机构知识库中博士论文、博士研究的原始数据文件以及硕士论文收藏量都比较大, 原因在于大学对这些内容强制提交的要求^[13]。日本机构知识库中纪要论文 (Kiyō) 则是主要的资源, 接下来是学术期刊论文 (见表 2)。

Kiyō 在日本学术界非常流行, 它是一种包括许多研究论文和报告的内部期刊, 可以被看作发掘研究人员的摇篮以及一些探索性研究的发源地^[14]。许多日本机构包括医院和企业等都发行纪要论文期刊, 但其中以教育机构居多。Kiyō 的主要特点就是机构内部的任何人都可以在该期刊上发表文章, 而且机构还鼓励内部成员在其上发表文章。Kiyō 主要是在机构内出版部门成员之间流通, 只有很少的 Kiyō 期刊在机构外部流通。虽然 Kiyō 不易获取并

缺乏质量控制, 但其已经成为日本人文和社会科学出版物的主体^[15]。根据国际机构知识库调查结果, 美国机构知识库中大约有 742 篇非学术审查的文章 (non - refereed publications)^[16]。McDowell 对美国机构知识库的研究发现学者的灰色文献 (即没有经过同行评议但实际上具有学术性的文献) 只占其 IR 内容的 23 %^[17]。虽然 Kiyoo 不是唯一的灰色文献类型, 但其已占有所有内容的 40.2 %。在日本已建成的 IR 中, Kiyoo 都是主要的资源, 即使有一些 IR 的资源类型较少, 但仍然以 Kiyoo 为主^[18]。这些表明了日本在学术出版方面的不同, 而机构知识库正好弥补这些差异, 使更多的学术读者可以获取 Kiyoo 电子期刊, 这可能会有助于高质量的学术产出。

表 2 机构知识库的资源类型

资源类型	数量	所占比例	资源类型	数量	所占比例
期刊论文	182 954	27.16 %	学位论文	37 971	5.64 %
纪要论文	270 851	40.21 %	会议论文	43 174	6.41 %
演示报告	1 700	0.25 %	图书	15 334	2.28 %
技术报告	3 060	0.45 %	研究报告	11 133	1.65 %
文集	26 773	3.97 %	预印本	254	0.04 %
学习资料	3 990	0.59 %	数据或数据集	468	0.07 %
软件	8	0.00 %	其他	75 983	11.28 %

(数据来源: 根据 JAIRO 网站整理。)

2.3 各类资源的语言分布

就国际范围内的用户而言, 他们比较关心各种语言与英语之间的切换, 而在 JAIRO 中大约 59 % 的机构知识库提供了日语和英语两种界面, 便于其他国家的学者进行浏览。此外, 机构知识库内容的语言可能会反应出机构出版模式的差异^[19]。虽然日本 IR 内容的语言有 31 种^[20], 但是主要以英语和日语为主。从下图我们可以看出, 期刊论文中英文文献所占的比例要大于 Kiyoo 中英文文献所占的比例, 这说明 Kiyoo 面向的对象主要是机构内部人员。

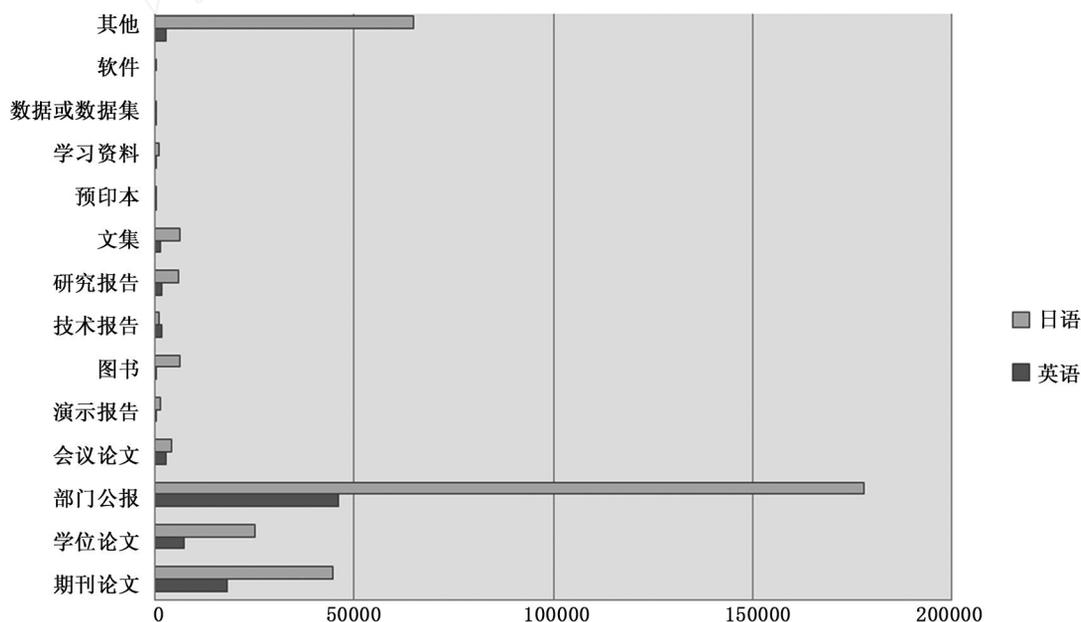


图 1 各类资源语言分布情况

(数据来源: 根据 http://irdb.nii.ac.jp/analysis/index_e.php [2009/05] 网站上的信息整理所得。)

虽然大部分的资源里都有英文资料, 但是在期刊论文、技术报告和预印本这 3 种资源中英文资料所占的比例

较大, 分别是 28.7%, 38.2%, 58.9%。这些形式的学术成果在世界其他地方是众所周知的, 也是普遍的学术共享与交流方式。既然期刊论文和其他一些资源的英文版本是其主要特征之一, 那么就表明这些资源有不同的目标^[21]。虽然这些英文资料对日本的学生和其他人员造成了使用方面的障碍, 但是它们是面向国际研究人员的。这样即使每个研究机构内英文资料的使用者比较少, 但它们在向全球展现日本的机构知识库, 也就使日本的学术研究在更广的范围内实现共享, 从而提高日本大学和研究机构的国际影响力和知名度。

3 对我国 IR 发展的启示

我国的机构知识库还处在起步阶段, 其数量和规模都不大。根据 OpenDOAR 的统计, 我国目前有 6 所机构知识库 (香港 4 所, 内地 2 所), 其中厦门大学和中国科学院国家科学图书馆是实用型 IR, 清华大学、北京理工大学和浙江大学是试验型 IR, 其他一些高校更多的是处于关注或计划建设阶段。这些 IR 的建设大都以 DSpace 系统为原型, 其收藏的内容和应用范围还比较分散。

3.1 认识机构知识库的价值

要加快机构知识库的建设, 需要让不同层面了解并认识到机构知识库的重要性, 这样才能引起其对机构知识库的关注, 才能从根本上提高其积极性。高校等应该通过各种渠道积极宣传机构知识库建设的重要性, 包括国家科研机构或学术团体组织丰富多样的研讨会, 使机构与学者都认识到机构知识库对其自身发展的利益, 提高机构与学者对机构知识库的认识, 提高他们对开放存取理念的认同度。我国目前对机构知识库认同方面仅限于图书情报领域的学者或实践者, 其他领域的学者则了解不多。所以要提高对机构知识库价值的认识, 需要国家从宏观层面进行调控, 就像日本文部科学省的 NII 一样积极推动机构知识库的建设。

3.2 加强国家层面的支持

根据上文, 日本的机构知识库的数量及其内容一直处于增长势头, 这在很大程度上得益于日本政府的支持及日本国家情报机构 (NII) 的倡导。所以, 国家层面的政策支持也是我国机构知识库建设发展最需要的推动力之一。比如作为我国权威性的国家自然科学基金会、国家社会科学基金会、教育部项目基金会以及其他各级政府资助委员会, 可以建议受其资助的科学研究成果纳入到公开获取的机构知识库当中^[22]。同时, 我国也可以借助日本的经验, 由专门的机构负责协调各个高校和研究机构, 促进他们之间的合作, 进行项目分配, 还可以建立专门的网站便于学者进行跨部门搜索及相关分析。此外, 国家层面还应该给予资金支持, 因为对于机构来说, 运转资金是有限的, 让他们再投入较多的资金支持图书馆或其他相关机构建立知识库确有困难, 所以国家更应该从宏观政策方面进行引导, 提供支持。

3.3 扩充机构知识库的内容

首先, 从机构知识库的语言界面来看, 除了香港的 4 所大学是英文界面, 内地的机构知识库只提供中文界面, 没有可供选择的英文界面。这样既不利于国际上的学术读者了解我国的学术研究成果及现状, 也不利于提高我国学术机构及高等院校的国际影响力和知名度。从日本机构知识库的访问情况来看, 美国、英国、德国等非日语国家对其有较多访问, 这与其将近 59% 的机构知识库提供英文界面是分不开的。所以我国机构知识库也应该建立英文界面, 便于国际学者的访问。其次, 从提供的内容类型上来看, 目前内容比较全面的是中国科学院国家科学图书馆机构知识库, 其主要有 8 种资源类型, 分别是期刊论文、会议论文、学位论文、专著、文集、研究报告、演示报告以及其他, 而且每种类型的资源都是可获取的。厦门大学的学术典藏库和香港几所大学的机构知识库则主要是会议论文、期刊论文和演示报告这几个方面, 清华大学和浙江大学的资源类型则更少。所以, 现阶段应将机构知识库的重心放在知识库内容的扩充上。各机构应采取各种措施进行广泛的宣传, 吸引内部成员将研究资料存放在机构知识库中。当然, 内容项目应该优先考虑展示机构的强项和成果, 包括国家资助的研究论文集及各种合作项目。此外, 我国的机构知识库也应该致力于灰色文献以及有关的数据集、软件等的推广和数字化, 比如很难搜索到的机构内部学术文章和论文, 从而充分发挥机构知识库的学术价值。

注释

[1] Lynch, C. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *Portal: Libraries and the Academy*, 3 (2), 327 - 336

[2] 柯平, 王颖洁. 机构知识库的发展研究. *图书馆论坛*, 2006 (12), 243 - 248

[3] [4] <http://www.opendoar.org/>, 2009 - 05 - 30

(下转第 90 页)

DOAJ 为图书馆员和科研人员搜寻和访问这些期刊提供了一个高质量的工具。

8.4 Metacritic (<http://www.metacritic.com/>)

该网站是个收录了个人对电影、视频、电视、音乐、书籍和游戏等评价的可检索型数据库, 可以通过作品名称、类型、人物、元数值 (每个人对作品评论数值的加权平均值) 等途径进行检索。它在计算加权平均值时考虑到了个人评论和出版物本身的威望, 其中绿色、黄色、红色标示的元数值分别表示该作品很受欢迎、一般、不受欢迎, 所以通过它可以非常容易地确定哪些是最受欢迎的作品。该网站所收录的每件作品的信息包括内容提要, 该作品的创作者、出版者、发行者、作品评价等等。

8.5 The Library of Congress (LOC) (<http://www.loc.gov>)

该网站除了提供图书馆的馆藏目录和世界范围内图书馆所收藏的图书和资源的书目信息外, 还是美国国会图书馆所收藏的独特的数字收藏品和网络资源目录的门户。从它的主页可以存取到的重要资源包括美国记忆、历史收藏品展览、通向多元文化的全球门户、提供立法信息的托马斯等。该信息量丰富、组织良好的网站主页还展示了诸如历史上的今天、图书馆网络直播和值得关注的图书馆新闻及事件等方面的专栏和

信息。

8.6 WorldCat (<http://worldcat.org>)

WorldCat 是世界上最大的提供图书馆内容和服务的网络, 它通过参加 OCLC 全球合作的图书馆资料库建造并维持, 目前通过它强大的目录体系可以从世界各地一万多所图书馆中检索出超过 10 亿种的资料。用户可以在自己的网站上添加一个 WorldCat 搜索框, 下载浏览器插件便不仅能够从浏览器工具栏中检索当地图书馆中的资源, 而且能够从全世界图书馆中检索自己所需要的资源。此外, 用户还可以免费在 WorldCat 网站上注册, 建立公共的和不对外的图书馆资料列表, 在所喜爱的特定主题或者类型的图书和电影上加注或为研究项目组织资源等, 可以建立公共的资料档, 便于与其他 WorldCat 用户共享兴趣等。

参考文献

- 1 <http://www.ala.org/ala/mgrps/divs/rusa/sections/mars/marspubs/marsbestindex.cfm>, 2009 - 11 - 03
- 2 <http://www.ala.org/ala/mgrps/divs/rusa/sections/mars/marspubs/marsbestfreewebsites/marsbestfree2009.cfm>, 2009 - 11 - 03

陈丽君 许昌学院图书馆。

(上接封三)

- [5] [8] Murakami Y, Adachi J. Digital Libraries: Achievements, Challenges and opportunities. Springer Berlin, 2006: 540 - 549
- [6] Murakami Y, Tutiya S, Sato Y. The interim evaluation method of the national project for institutional repositories in Japan. IATUL Proceedings, 2007: 1 - 8
- [7] Katherine K Matsuura. Japan's Institutional Repositories: Where Did They Come From and Where Are They Headed?. University of North Carolina, 2008
- [9] [11] [16] [18] [19] [20] [21] http://jairo.nii.ac.jp/help/en/about_us.html, 2009 - 05 - 25
- [10] <http://jairo.nii.ac.jp/stats/en/index.do>, 2009 - 05 - 30
- [12] <http://www.nii.ac.jp/irp/en/list/>, 2009 - 06
- [13] 何林. 美国机构知识库发展现状对我国发展机构知识库的启示. 图书馆论坛, 2008 (6), 101 - 103
- [14] K Keiko Hisama. Culture and Nursing Research: An Empirical Analysis of Kyo Journals. Bulletin of allied medical sciences Kobe : BAMS (Kobe), 1997 (13): 85 - 94
- [15] Kamada, Hitoshi. Portal: Libraries and the Academy. Kyo Journals and Scholarly Communication in Japan. 2007 (7): 375 - 383
- [17] Cat S McDowell. Evaluating Institutional Repository Deployment in American Academe Since Early 2005. D - Lib Magazine, 2007 (10)
- [22] 邓君. 机构知识库建设模式与运行机制研究. 吉林大学, 2008

王艳霞 中国科学院国家科学图书馆 2007 级硕士研究生。

叶小梁 中国科学院国家科学图书馆研究员。