

机构知识库联盟发展现状及关键问题分析

曾 苏^{1,2} 马建霞¹ 祝忠明¹

¹ (中国科学院国家科学图书馆兰州分馆/中国科学院资源环境科学信息中心 兰州 730000)

² (中国科学院研究生院 北京 100190)

【摘要】在阐述机构知识库联盟概念、建设意义的基础上,对国内外主要国家层面和区域性机构知识库联盟项目的发展现状进行了总结,并从机构知识库联盟可持续发展的角度提出5个关键问题:组织管理模式、政策框架、知识产权策略、技术组织方式及服务,以期为机构知识库联盟在我国的发展提供参考。

【关键词】机构知识库;机构知识库联盟;发展现状;关键问题

【分类号】G250

Analysis of the Development and Key Issues of Institutional Repository Network

Zeng Su^{1,2} Ma Jianxia¹ Zhu Zhongming¹

(The Lanzhou Branch of the National Science Library / The Scientific Information Center for Resources and Environment, Chinese Academy of Sciences, Lanzhou 730000, China)

(Graduate University of Chinese Academy of Sciences, Beijing 100190, China)

【Abstract】Based on the analysis of the concept and significance of institutional repository network, the article summarizes the development of major projects of national and regional institutional repository network. In order to provide reference to China, the authors point out key issues of the development of institutional repository network from the perspective of sustainable development: organization model, policy framework, intellectual property strategy, technology issues and services.

【Keywords】institutional repository; institutional repository network; development analysis; key issues

在开放获取理念兴起的背景下,机构知识库(Institutional Repository,以下简称IR)的建设逐步成为国内外研究机构和大学的热点。2005年,美国网络信息联盟(CNI)对其成员(121所大学和81所文科学院)的IR部署情况开展了调查研究,结果显示:40%的大学、6%的学院已经运行IR,在没有建立IR的机构中88%的大学和21%的学院计划加入到IR联盟系统中^[1]。根据现有的资料,大部分研究机构和大学的IR中收录的文献记录都少于1000条,而一些小型机构通常没有资源、资金来建立或支持IR的正常运行。由于资金和内容征集等各方面因素的制约,单个机构IR的建设和正常运行受到一定的影响,IR联盟是未来机构知识库发展的趋势之一。本文对IR联盟的概念进行了界定,并在此基础上进行了IR联盟建设意义、发展现状和关键问题分析,以期为国内IR联盟发展提供参考。

1 机构知识库联盟概念

国内外学者对IR联盟还没有统一的定义,比较典型的主要包括:(1)美国网络信息联盟执行总裁Clifford A.Lynch将其定义为:机构知识库之间实现跨库检索和内容交换,实现数据备份、保存、故障恢复及其他功能,促进跨机构的学术交流和合作^[2]。(2)CrossRef.org:一些机构构建的共享知识库或多个互操作机构知识库,为联盟机构成员提供存档和访问服务^[3]。(3)李广建等:学术机构通过合作的方式,将各自的资源库整合起来,统一提供数字化服务,支持不同IR之间的数据交换和共享,支持跨库无缝检索,同时各分布式IR拥有数据备份、保存及故障恢复能力^[4]。(4)邓君在其博士学位论文中将IR

联盟定义为：两个以上机构联合构建机构知识库，通过合作的方式，实行机构间资源共享，统一提供知识传播与知识服务，机构知识库联盟模式可以分为集中式与采集式^[5]。(5) 王文华：机构联盟知识库，就是指几个机构以一个机构为基地联合构建知识库，通过合作的方式，将各自的资源库整合起来，统一提供数字化服务^[6]。

笔者在比较以上定义的基础上，将机构知识库联盟界定为：两个以上大学、研究机构及相关组织通过合作的方式构建机构知识库或共享机构知识库资源，以集中存缴、元数据收割等方式统一提供知识传播和知识服务，以实现不同机构间知识产出的共享、利用。IR 联盟按地理位置划分为：校（所）际 IR 联盟、地区级 IR 联盟、国家级 IR 联盟、洲级 IR 联盟；按技术组织模式可划分为：集中式 IR 联盟、采集式 IR 联盟、混合式 IR 联盟。

2 机构知识库联盟建设意义

2.1 更大范围内的知识产出保存、共享

IR 是在开放获取的背景下逐步发展起来的，但目前并不是所有的 IR 都是完全开放获取的，每个机构和大学的 IR 都有一定的访问控制策略。IR 联盟的发展，使得联盟内机构成员的知识产出得到统一保存、利用和共享，并通过统一的界面提供服务，有利于突破机构的限制共享知识。国家级 IR 联盟、洲级 IR 联盟的成熟和发展，对实现学术信息资源的长期保存和开放获取意义深远。

2.2 节约成本，体现“规模效益”

无论采取哪种技术模式，IR 联盟的建设可以实现构建成本和运营成本的“规模效益”。集中式 IR 联盟由多个机构、大学共同构建和维护一个系统，系统建设和运营成本实行分摊，使得单个机构构建 IR 成本大大降低，小型机构加入已有的集中式 IR 联盟可实现成本节约和资源扩充的双赢效果。采集式 IR 联盟成员一般选择遵循相同标准化开放接口的软件平台，联盟成员在软件安装、运行、维护、管理及人员部署方面分享经验，并且共同解决 IR 建设过程中的技术问题，这可实现 IR 建设过程中的成本节约。

2.3 促进单个机构 IR 的建设推广

根据 ARROW、DAREnet 等国家级 IR 联盟项目的实践经验，IR 联盟的建设可在一定程度上推动单个机构 IR 构建、部署。国家级 IR 联盟的构建，国家制定相关政策、提供资金支持，可实现全国范围内 IR 数量的增长和顺利部署。由于资金、技术力量缺乏，小型机构无力独自构建 IR 服务系统，通过参与区域性联盟方式实现本机构 IR 的顺利部署，这也将促进单个机构 IR 数量的增长。

2.4 构建全国知识基础保存设施的重要步骤

目前，国内 IR 建设还处于起步时期，仅有少数研究机构和大学建设了 IR 系统。国内科研机构和研究人员每年要产生大量的科研成果，无论在机构层面还是国家层面都需要构建 IR 一类的知识产出保存系统。作为科研信息环境基础设施的 IR，在科研产出集中管理、长期保存、共享、利用方面发挥着重要作用，因此可通过构建不同层次的 IR 联盟，继而推动全国性 IR 联盟，实现对国家科研产出的长期保存和有效管理，并且作为全国知识基础保存设施对科研成果进行集中呈现和统一服务。通过分层构建

IR 联盟的方式, 逐步构建上级联盟, 进而发展成为全国性知识基础保存设施。

3 机构知识库联盟发展现状

随着开放获取运动的不断发展, 国外 IR 建设在理论和实践方面取得了很大发展。在德国、荷兰、挪威、澳大利亚, IR 已经成为学术机构的普遍基础设施, 美国、法国、日本等国 IR 发展也较快。随着 IR 理论研究和实践的不断推进, 荷兰、法国、澳大利亚、美国等国及欧洲已开始构建 IR 联盟, 在地区、国家甚至欧洲范围进行知识成果集中存缴和呈现。国内 IR 建设还处于起步阶段, 仅有少数研究项目对 IR 联盟进行了探索试验。

3.1 国外 IR 联盟发展现状

3.1.1 主要国家层面 IR 联盟

国家层面 IR 联盟主要包括: 法国 HAL、荷兰 DAREnet、澳大利亚 ARROW、日本 JAIRO、德国 OA-Network、英国 JISC RepositoryNet、欧盟 DRIVER^{[7]-[13]}。

表 1 国家层面 IR 联盟项目

	参与机构	技术模式	收录文献类型	条目数量	服务
HAL	CNRS、INSERM、INRA、86 所大学	集中式	期刊论文、会议论文、图书、学位论文、研究报告等	122 540	集中检索、浏览; 向主题仓储提交内容; 生成学科门户; 使用统计; 出版物清单等
DAREnet	所有大学和部分研究机构	采集式	学术出版物、学位论文	224 257	集中检索、浏览; 专题检索、浏览
ARROW	24 所大学	采集式	e-prints、学位论文、电子出版物	284 501	集中检索、浏览;
JAIRO	98 所大学和研究机构	采集式	期刊论文、学位论文、会议论文、演示文档、图书、技术报告、研究报告等 14 种类型	637 237	集中检索、浏览
OA-Network	拥有 DINI 证书的 IR	——	——	——	未实现统一检索、浏览;
JISC RepositoryNet	所有大学	——	——	——	未实现统一检索、浏览; Intute Repository Search
DRIVER	欧洲国家层面 IR 联盟	混合式	期刊论文、会议论文、图书、学位论文、研究报告、报纸等	1 000 000	集中检索、浏览;

注: a) 该表数据统计截止 2009 年 5 月 18 日; b) “——” 表示表示该项暂无统计

表1显示, HAL、DAREnet、ARROW、JAIRO、DRIVER项目通过集中提交或元数据收割的方式实现了本国(洲)知识产出的聚合和检索;德国OA-Network、英国JISC Repository还未实现本国知识产出的集中呈现和检索,将随着项目的深入得以实现。国家层面IR联盟发展呈现以下特点:①国家层面IR联盟出现在IR发展迅速的国家,其本国范围内已经有相当数量的IR,为IR联盟的构建提供了有利条件;②国家层面IR联盟大都由一些著名的大学、研究机构牵头,以项目的形式构建IR联盟,不断吸收其他机构加入联盟中,从而实现本国范围内知识产出的集中呈现和长期保存;③国家层面IR联盟的构建不仅仅是技术系统或平台的部署,更是推动国家层面知识产出有效管理、长期保存的战略实践。

3.1.2 主要区域性IR联盟

区域性IR联盟主要有:英国SHERPA项目资助的SHERPA—LEAP联盟、白玫瑰知识库联盟^{[14]-[15]};美国华盛顿研究图书馆联盟构建的ALADIN联盟、犹他知识库联盟、科罗拉多数字知识库联盟、德州数字知识库联盟、俄亥俄州数字知识库联盟、佐治亚知识库联盟^{[16]-[22]}。

表2 区域性IR联盟项目

	参与机构	技术模式	收录文献类型	条目数量	服务
SHERPA—LEAP 联盟	伦敦大学 13 个学院	——	——	——	未实现集中呈现
白玫瑰知识库联盟	利兹大学、设菲尔德大学、约克角大学	集中式	研究论文、会议论文、专著、图书、专利等	5 240	统一检索和浏览
ALADIN 联盟	华盛顿研究图书馆联盟 7 个成员机构	集中式	研究论文、学位论文、多媒体资料、数据等	4 147	统一检索和浏览
犹他知识库联盟	犹他研究图书馆联盟的 5 所大学	集中式	研究论文、学位论文、本科生荣誉方案、工作文件、研究报告等	4 713	统一检索和浏览
科罗拉多数字知识库联盟	科罗拉多研究图书馆联盟 11 个成员机构	采集式	研究论文、多媒体资料等	——	统一检索和浏览
HELIN 机构知识库联盟	HELIN 图书馆联盟成员机构	采集式	研究论文、学位论文、会议论文等	789	统一检索和浏览
德州数字知识库联盟	德州 5 个研究图书馆协会成员大学	采集式	期刊论文、学位论文、会议论文、研究报告等	7 336	统一检索和浏览
俄亥俄州数字知识库联盟	俄亥俄州 11 所大学和文科学院	采集式	学位论文、课件、多媒体资料等	168 520	统一检索和浏览
佐治亚知识库联盟	乔治亚州的大学和研究	采集式	学位论文、研究报告、期刊论文、研究数据等	218	统一检索和浏览

注: a) 该表数据统计截止 2009 年 5 月 18 日; b) “——” 表示该项暂无统计

区域性IR联盟发展特点主要有:①区域性IR联盟的构建,大都是在原有图书馆联盟基础上构建IR联盟。图书馆联盟合作基础好、发展时间长,为IR联盟的构建提供了较为有利的条件。②区域性IR联盟的构建,在一定程度上解决了小型机构IR建设资金和内容征集方面的制约,推动了单个机构IR的

建设和推广。

3.2 国内 IR 联盟发展现状

国内IR建设还处于起步阶段,仅有香港大学、香港科技大学、香港城市大学、香港理工大学、清华大学、厦门大学、浙江大学、中国科学院力学研究所、中国科学院国家科学图书馆构建了IR系统^[23]。国内IR研究和实践落后于欧美等国,对IR联盟的研究还处于探索试验阶段,相关研究主要有:

中国科学院国家科学图书馆的“全院联合机构仓储体系建设”项目提出了构建全院联合机构仓储体系的构想:即在以研究所为单元构建各所IR系统的基础上,通过元数据开放获取和内容聚合的方式,建立起全院联合机构仓储服务体系^[24]。该项目还处于分散建设阶段,仅对中科院力学研究所、中国科学院图书馆进行了IR部署,并已经在中科院西安光机所、软件所、半导体所等启动实施,但实施单位数量、尚不成规模。

中国香港地区的HKIR^[25](Hong Kong Institutional Repositories)是由香港科技大学图书馆开发的演示系统,实现对大学教育资助委员会所资助科研单位(8所大学和研究所)知识产出的聚合和联合搜索,截止到2009年5月共收割文献127743篇。

中国台湾地区的“建置机构学术成果典藏计划”采用“分散建置、集中呈现”的原则:截止到2009年5月,共有82所大学和研究机构使用台湾大学的NTUR系统建立本机构的典藏系统,并以TAIR^[26]入口网站作为台湾学术研究成果累积、展示与利用的窗口^[27]。

4 机构知识库联盟发展关键问题分析

4.1 组织管理模式

IR联盟的组织管理模式,是构建IR联盟的基础和先决条件。从IR联盟总体组织建设和推进的角度看,IR联盟可分为“自上而下”和“自下而上”两种组织模式。自上而下的组织模式,即相关大学或研究机构通过合作的方式构建IR联盟,共同参与IR规划、政策制定、内容管理、系统管理及服务功能设计,共享技术经验、软硬件和实行成本分摊,实现本机构IR的部署和实施;自下而上的组织模式,即相关研究机构或大学在各自运营和维护本机构IR服务系统的基础上,以资源扩充、共享为目的,以集中提交或元数据收割方式实现机构知识产出的聚合和集中呈现,以联盟系统的方式提供服务。

目前,国外IR联盟主流构建模式为自上而下模式。这种模式被广泛采用的原因主要体现在:首先,国家层面IR联盟的构建是一项复杂、艰巨的任务,只有在国家政策和资金支持下,自上而下推动全国范围内研究机构和大学IR系统构建,才能实现全国范围内知识资源的集中共享和揭示;其次,现有的区域性IR联盟大都在原有图书馆联盟的基础上构建的,联盟有较好的合作基础,这有利于联盟成员分工协作进行IR联盟构建,从而实现联盟成员单个IR的部署;第三,由于资金和内容征集的原因,小型机构和大学无力构建自己的IR服务系统,通过合作构建IR联盟或直接加入现有IR联盟的方式实现本单位知识资源的保存和共享。

4.2 政策框架

IR联盟发展政策是构建IR联盟不可或缺的观念、原则和行动指南,为其发展提供重要依据和宏观指导,是IR联盟可持续发展的关键因素和重要保证。IR联盟构建过程中涉及的政策主要包括:组织管理机制、资源管理政策、权利管理政策、长期保存政策。

表3 IR联盟政策体系与内容

政策	具体内容
组织管理机制	构建目标、组织形式、职责分工、人员配置、管理机制、经费支持与分摊
资源管理政策	资源聚合政策、质量控制政策、所有权政策、揭示政策、访问政策
权利管理政策	软件平台版权、合理使用、隐私权
长期保存政策	元数据描述、文档格式、撤销政策、数据保存和备份

4.3 知识产权策略

由于担心与出版商、资助机构的潜在冲突,研究人员不愿意提交或极少数提交科研成果到IR联盟服务系统中,因此IR联盟发展过程中应采取有效的知识产权策略。可采取的知识产权策略包括:以机构的名义,和国内外出版商进行协商和签订协议,争取作者自存储的权利并可向公众免费提供接入、检索、浏览、下载服务;协调好作者、资助机构、出版商三方的关系,区别对待研究人员知识产出的各种情况,避免出现知识产权纠纷;IR联盟应设立相关委员会或工作人员,负责解答和处理研究人员自存储过程中知识产权方面的疑惑;充分利用诸如SHERPA/RoMEO^[28]之类的版权查询系统,并将其嵌入到IR提交页面中,方便研究人员提交知识产出过程中查询相关信息;IR联盟服务系统中,应包含指向出版商主页的链接,并提供正确的引用信息,体现出版商公开发表出版权利;加强对创作共用协议(Creative Commons^[29])的宣传,鼓励研究人员接受并积极利用该协议,灵活运用著作权并保留相应权利。

4.4 技术组织方式

IR联盟不仅要实现机构内部数字知识产出的长期保存和有效管理,还要实现联盟机构内资源的互操作和共享,这就需要稳定技术平台和相关技术的支撑。IR联盟服务系统应根据需要选择合适的技术组织方式和技术方案,以达成联盟构建目标。IR联盟技术组织模式主要分为:集中式IR联盟指多个机构建立单一、集中的IR服务系统,联盟机构成员直接将元数据和内容提交到集中服务器中,对数字资产实现统一保存、利用、传播;采集式IR联盟指联盟成员机构分别建立各自的IR服务系统,通过元数据收割的方式实现联盟机构资源的集中呈现和揭示,提供统一的检索入口和界面,数字资源仍分布在各机构IR中;混合式IR联盟,是集中式IR联盟和采集式IR联盟的结合,既通过集中提交方式聚合资源,又通过采集的方式收割元数据。IR联盟服务系统涉及相关技术包括:数字对象标示符、作者标示符、复合数字对象描述、元数据收割、长期保存技术、单点登录技术等。

4.5 服务

IR联盟提供的服务,对提高IR联盟知名度、长期运营、内容部署起到十分重要的作用,是影响IR联盟发展的关键因素之一。IR联盟可提供的服务主要包括以下方面:

- (1) 一般服务:统一浏览、统一检索、简单快速提交、数字化、格式转换、RSS订阅、使用统计等;
- (2) 增值服务:出版物清单自动生成、数字资源导出服务、主题门户生成、研究人员及研究团体门户网站生成服务、单个机构IR门户生成服务、引文索引服务、支持科研评价等;
- (3) 资源揭示服务:在全球性知识库门户(OAster、OpenDOAR、ROAR)登记注册、搜索引擎索引、提供与外部信息系统的接口等。

5 结语

本文在分析 IR 联盟概念、意义的基础上,对 IR 联盟的发展现状进行了阐述,并分析了 IR 联盟发展中的关键问题。我国 IR 建设还处于起步阶段,大量研究机构和大学已经或将要进行 IR 系统服务建设。由于资金、技术及内容征集方面因素的制约,并非所有机构愿意建立本机构的 IR 服务系统。相关机构合作建立 IR 联盟或加入已有的 IR 联盟,将实现资本节约、成本分摊、经验共享,对我国 IR 建设和快速增长有着重要促进作用。探索适合中国实际情况 IR 联盟发展模式,是下一步研究的重要内容。

参考文献

- [1]Clifford A. Lynch, Joan K. Lippincott. Institutional Repository Deployment in the United States as of Early 2005[EB/OL]. [2009-2-10].
<http://www.dlib.org/dlib/september05/lynch/09lynch.html>
- [2]Clifford A. Lynch. Institutional Repositories: Essential Infrastructure for Scholarship in the DigitalAge[EB/OL]. [2009-2-10].
<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>.
- [3]CrossRef Glossary Version 1.0[EB/OL]. [2009-3-10].
https://www.policypress.org.uk/images/upload/pages/cr_glossary_1.pdf
- [4]李广建,黄永文,张丽. IR:现代体系结构与发展趋势[J]. 情报学报,2006,25(02):236-241
- [5]邓君. 机构知识库建设模式与运行机制研究[D]. 长春:吉林大学,2008
- [6]王文华. 知识库发展的新模式——机构联盟知识库[J]. 情报科学,2008,26(03):373-376
- [7]HAL[EB/OL]. [2009-2-27].<http://hal.archives-ouvertes.fr/index.php?langue=en&halsid=r9t66q17gls22mskasmipt09m5>
- [8]DAREnet[EB/OL]. [2009-2-27].<http://www.narcis.info/index/tab/narcis>
- [9]ARROW[EB/OL]. [2009-2-27].<http://arrow.edu.au/>
- [10]JAIRO[EB/OL]. [2009-5-18]. <http://jairo.nii.ac.jp/en/>
- [11]OA-Network[EB/OL]. [2009-2-27].http://www.dini.de/fileadmin/oa-netzwerk/PM_OA-Netzwerk_Projektstart_en_080116.pdf
- [12]JISC RepositoryNet[EB/OL]. [2009-2-27].
<http://www.jisc.ac.uk/publications/publications/repositorynet.aspx>
- [13] DRIVER[EB/OL]. [2009-2-27].<http://www.driver-repository.eu/>
- [14]SHERPA—LEAP[EB/OL]. [2009-2-27].<http://www.sherpa-leap.ac.uk/about.html>
- [15]WRCER[EB/OL]. [2009-2-27].<http://eprints.whiterose.ac.uk>
- [16]ALADIN[EB/OL]. [2009-2-27].<http://dspace.wrlc.org/>
- [17]Utah Academic Library Consortium[EB/OL]. [2009-2-27].
http://harvester.lib.utah.edu/utah_ir/index.php/index
- [18]Alliance Digital Repository[EB/OL]. [2009-2-27].
<http://adr.coalliance.org/adrlib/>
- [19]HELIN[EB/OL]. [2009-2-27].<http://helindigitalcommons.org/>
- [20]Texas Digital Library Repositories[EB/OL]. [2009-2-27]. <http://repositories.tdl.org/>
- [21]OhioLINK Digital Resource Commons[EB/OL]. [2009-2-27]. <http://drc.ohiolink.edu/>
- [22] GALILEO KNOWLEDGE REPOSITORY[EB/OL]. [2009-5-18]. <http://gkr.gatech.edu/>
- [23]曾苏,马建霞,汤天波,韩珂. 国内科研机构 and 高校机构知识库规划建设现状与问题研究[J]. 现代图书情报技术,2009(1):50-57
- [24]祝忠明,马建霞,张智雄,孙坦. 中国科学院联合机构仓储系统的开发与建设[J]. 图书情报工作,2008(9):90-93,144
- [25] Hong Kong Institutional Repositories[EB/OL]. [2009-5-30]. <http://lbapps.ust.hk/hkir/>
- [26] Taiwan Academic Institutional Repository[EB/OL]. [2009-2-10].
<http://tair.lib.ntu.edu.tw/>

- [27] 机构典藏计划网站 [EB/OL]. [2009-2-10]. <http://ir.lib.ntu.edu.tw/Wiki.jsp?page=Main>
[28] SHERPA/RoMEO [EB/OL]. [2009-2-10]. <http://www.sherpa.ac.uk/romeo/>
[29] Creative Commons [EB/OL]. [2009-2-27]. <http://creativecommons.org/>

作者简介:

曾苏, 男, 1986年生, 硕士研究生, 发表论文6篇

通讯地址: 甘肃省兰州市天水中路8号中国科学院国家科学图书馆兰州分馆研究生学习室 730000

Email: zengs@mail.las.ac.cn

马建霞, 女, 1972年生, 研究馆员, 硕士生导师, 发表论文20余篇

祝忠明, 男, 1968年生, 研究馆员, 硕士生导师, 发表论文30余篇