

北京大学图书馆研究支持服务探索与实践丛书

第2卷 (90) 目录附录并图

机构知识库构建指南

崔海媛 著

海洋出版社

2019年·北京

图书在版编目 (CIP) 数据

机构知识库构建指南：北京大学图书馆研究支持服务探索与实践/
崔海媛著. —北京：海洋出版社，2019. 4

ISBN 978-7-5210-0346-8

I. ①机… II. ①崔… III. ①图书馆工作-研究 IV. ①G25

中国版本图书馆 CIP 数据核字 (2019) 第 072391 号

丛书策划：高显刚

责任编辑：杨海萍 张欣

责任印制：赵麟苏

海洋出版社 出版发行

<http://www.oceanpress.com.cn>

北京市海淀区大慧寺路 8 号 邮编：100081

北京朝阳印刷厂有限责任公司印刷 新华书店发行所经销

2019 年 5 月第 1 版 2019 年 5 月北京第 1 次印刷

开本：787mm×1092mm 1/16 印张：26.5

字数：352 千字 定价：68.00 元

发行部：62132549 邮购部：68038093 总编室：62114335

海洋版图书印、装错误可随时退换

《机构知识库构建指南》

编委会

崔海媛 罗鹏程 李国俊

刘丹 吴越 徐清白

前言



自1998年“自由扩散科学成果运动”提出开放获取倡议后，二十年来，全球开放获取运动蓬勃发展，多个国家、资助机构、大学与科研机构制订开放获取政策，机构知识库发展迅速，已经成为全球学术交流中重要的学术资源。与其他国家对比，中国高校和科研机构无论在开放获取政策发布，还是机构知识库建设服务方面，都落后于世界。2018年11月，在开放获取知识库登记网站（Registry of Open Access Repositories, ROAR）登记在册的全球开放获取知识库数量为4 735个，中国大陆只有92个（中国科学院研究所知识库73个，其他机构19个），香港地区10个，台湾地区83个，建设的机构知识库规模和影响力都非常有限。调研显示，虽然中国高校图书馆参与机构知识库的积极性很高，但因政策力度、技术能力、人员配备和建设经验不足，难以有效开展与推动机构知识库建设。2017年开始，中国高校开始双一流建设，在教学和科研上，进入世界一流水平。机构知识库建设，应纳入各高校双一流建设规划中，中国机构知识库如果能在全球学术交流生态系统中发挥作用，有助于提升中国高校的学术影响力，支持双一流建设。

本书调研和分享了全球机构知识库的建设经验和成功案例，并基于编著者多年的研究和实践经验，对机构知识库的缘起、概念、发展历程、开放获

取政策制定、软件平台选择和构建、开源软件 Dspace 应用、元数据标准规范制定、内容建设方法、服务推广策略、机构知识库与科研管理系统的融合、新一代机构知识库发展趋势等做了全面介绍，力图为国内同行提供从理论到实践的全面建议和参考。本书编著者是北京大学图书馆信息化与数据中心技术团队成员，具有丰富的图书馆新技术研究和开发经验，从事多年机构知识库的研究和建设工 作，为北京大学机构库建设和发展做了多方面的贡献。在本书中，我们将研究成果和建设经验分享给读者，期待能给国内各高校图书馆机构知识库建设提供参考和借鉴，推动中国机构知识库发展。由于能力、学识不足，书中内容难免会有疏漏和不足，也诚意恳请业界同行指正。

本书的写作分工如下：

崔海媛（北京大学图书馆）：第一章、第二章、第三章、第七章、第八章、第九章，并负责全书的内容设计、组织、体例、审校和统稿；

罗鹏程（北京大学图书馆）：第三章、第四章、第八章；

李国俊（北京大学图书馆）：第三章、第四章；

刘丹（北京大学图书馆）：第五章、第六章；

吴越（北京大学图书馆）：第二章；

徐清白（北京大学图书馆）：第二章、第四章。

崔海媛

2018年11月

目 录



| | |
|------------------------------|------|
| 第一章 机构知识库发展概述 | (1) |
| 1.1 开放获取运动 | (1) |
| 1.2 机构知识库发展概述 | (8) |
| 1.3 国际机构知识库发展 | (14) |
| 1.4 中国机构知识库发展情况 | (32) |
| 1.5 结语 | (39) |
| 参考文献 | (39) |
| 第二章 开放获取政策框架与制定 | (45) |
| 2.1 国内外开放获取政策发展现状 | (46) |
| 2.2 开放获取政策的内容框架 | (55) |
| 2.3 开放获取政策制定流程 | (62) |
| 2.4 开放获取政策制定建议 | (75) |
| 2.5 结语 | (84) |
| 参考文献 | (84) |



| | |
|--------------------------------|-------|
| 第三章 机构知识库软件平台建设指南 | (91) |
| 3.1 全球机构知识库软件平台应用概况 | (91) |
| 3.2 机构知识库的关键技术发展 | (93) |
| 3.3 机构知识库开源软件介绍 | (115) |
| 参考文献 | (136) |
| 第四章 机构知识库开源软件应用 | (141) |
| 4.1 DSpace 安装 | (141) |
| 4.2 DSpace 使用 | (154) |
| 4.3 DSpace 开发 | (169) |
| 4.4 PKU IR 的开发与使用 | (184) |
| 4.5 CHAIR Local 的开发与使用 | (187) |
| 4.6 NSFC-OR 的开发与使用 | (190) |
| 4.7 结语 | (192) |
| 参考文献 | (192) |
| 第五章 机构知识库元数据标准规范 | (194) |
| 5.1 数字图书馆资源描述常用元数据标准 | (194) |
| 5.2 国内外机构知识库元数据标准规范 | (201) |
| 5.3 机构知识库元数据规范制订 | (205) |
| 5.4 北京大学机构库元数据规范 | (206) |
| 5.5 国家自然科学基金基础研究知识库元数据规范 | (213) |
| 参考文献 | (215) |

| | |
|---------------------------------|-------|
| 第六章 机构知识库内容建设 | (218) |
| 6.1 内容建设模式 | (218) |
| 6.2 元数据质量控制 | (233) |
| 参考文献 | (248) |
| 第七章 机构知识库服务推广 | (251) |
| 7.1 发展现状与挑战 | (251) |
| 7.2 机构知识库服务研究综述 | (253) |
| 7.3 机构知识库服务评价指标 | (255) |
| 7.4 机构知识库服务推广策略 | (260) |
| 7.5 北京大学机构知识库服务推广案例 | (274) |
| 7.6 结语 | (278) |
| 参考文献 | (279) |
| 第八章 机构知识库与科研管理系统融合 | (281) |
| 8.1 科研管理系统 | (282) |
| 8.2 机构知识库与科研管理系统融合 | (299) |
| 8.3 北京大学融合建设案例 | (310) |
| 8.4 结语 | (314) |
| 参考文献 | (315) |
| 第九章 新一代机构知识库与学术交流生态系统发展趋势 | (319) |
| 9.1 新一代机构知识库研究综述 | (320) |
| 9.2 新一代机构知识库技术和功能特点 | (322) |
| 9.3 新一代学术交流生态系统 | (332) |

| | |
|--|--------------|
| 9.4 结语 | (345) |
| 参考文献 | (345) |
| 附录 1 国内外外部高校和资助机构开放获取政策 | (349) |
| 1 美国哈佛大学开放获取政策网站首页 | (349) |
| 2 美国哈佛大学文理学院开放获取政策 | (351) |
| 3 美国哈佛大学开放获取政策模板(2013 年版) | (352) |
| 4 美国哈佛大学开放获取政策常见问题 | (358) |
| 5 美国加利福尼亚大学校级开放获取政策 | (361) |
| 6 美国加利福尼亚大学开放获取政策常见问题 | (369) |
| 7 英国研究理事会开放获取政策 | (376) |
| 8 英国研究理事会开放获取政策常见问题 | (392) |
| 9 北京大学机构知识库开放获取政策(试行) | (400) |
| 10 国家自然科学基金会关于受资助项目科研论文实行开放获取的政策 声明 | (401) |
| 11 国家自然科学基金委员会基础研究知识库开放获取政策实施细则 | (402) |
| 附录 2 中国机构知识库用户调查问卷 | (406) |

第一章 机构知识库发展概述



1.1 开放获取运动

1.1.1 背景

学术交流是科学研究的重要组成部分，是科学研究的本质^[1]。自最早的两份学术期刊《学者周刊》和《哲学汇刊》于1665年问世以来，学术期刊一直是学术交流最为重要的正式渠道^[2]。随着信息技术的发展和进步，尤其是互联网的出现和迅速普及，网络出版与电子数据库成为学术出版趋势。三百多年来，出版者、数据库商与图书馆在学术交流活动中发挥着重要作用，但自20世纪70年代产生并愈发严重的“学术期刊危机”问题使得出版者与图书馆出现了矛盾。访问学术期刊和获取学术成果方面的“价格危机”（pricing crisis）和“许可危机”（permission crisis）^[3]，造成学术交流中，科研人员一方面不能有效快速地获取他人的研究成果，另一方面也不能无障碍传播研究成果。学术期刊的大幅度涨价，出版的商业化和兼并行为，使学术期刊的出版日益集中于少数商业出版者手中，学术期刊价格持续的高涨大大超出了图书馆文献采购经费预算的涨幅，迫使图书馆不得不削减期刊订购，造成价格

危机。电子期刊与数据库普及后，出版商设置访问限制，将图书馆订购的电子期刊使用许可授权限制在一定范围内，造成许可危机，阻碍了学术研究成果的广泛传播和有效利用。

为了解决危机，打破传播壁垒，利用信息技术促进学术交流，学术界、图书馆界和出版界不断探索新的学术交流方式。20世纪90年代末兴起与快速发展的开放获取运动（Open Access, OA），变革与推动了信息自由共享与传播的学术交流机制。不同于传统基于订阅的学术出版模式，开放获取倡导通过互联网进行学术成果无障碍传播，任何人可以在任何地点和任何时间，不受经济状态影响，平等获取并使用学术信息和研究成果^[4]。

1.1.2 开放获取定义

目前，开放获取的定义，较为全面且广泛使用的有两个：

2001年布达佩斯会议中将开放获取描述为：可以免费在互联网上得到，允许读者阅读、下载、复制、传播、打印、检索、获取文章全文，可以制作索引，把它们作为软件数据库，或者为了其他合法的目的使用而不存在资金、法律、技术障碍。对复制与传播的唯一约束就是，应确保作者有权利控制他们成果的完整性并使成果得到认可与引用^[5]。

2003年4月11日，在美国马里兰州召开的 Bethesda 会议上，定义开放获取具有两大特征^[6]：①保留部分权利。作者及著作所有权人授权所有使用者免费、不能撤回、全球性、永久的取用权利，只要恰当的署名原著者，就可以不拘理由以任何数字媒体形式，公开复制、使用、散布、传送及展示原作品，制作及散布衍生作品；印本形式只限少量及个人用途。②公开发布。包括所有附件、以上著作权声明在内的完整作品，以适当的标准档案格式存储在至少一个在线知识库中。该知识库必须由学术单位、学会、政府机构或其他声誉卓著的机构支持，采用开放、传播无限制、跨平台及长期保存政策。

以美国国立医学图书馆（NLM）与美国国立卫生研究院（NIH）联合建立的生命科学期刊全文数据库（PubMed Central）为例，PubMed 就是生物医学领域里符合上述要求的一个知识库。

维基百科对开放获取的定义是^[7]：开放获取是在网络范围内任何人都可以免费的、及时的、永久的、全文的联机获取数字科技与学术资料，主要是在同行评议的期刊上发表的研究资料。开放获取意味着任何用户在任何地方，都可以利用互联网链接、阅读、下载、存储、打印、使用和数据挖掘这篇文章的数字内容。开放获取的文章通常具有有限的版权和许可限制。简而言之，开放获取就是任何人在网络范围内都可以免费、及时、永久获取研究文章的全文。

美国研究图书馆学会（Association of Research Libraries, ARL）基于布达佩斯和 Bethesda 会议定义，将 OA 定义解释为：“开放获取”是在基于订阅的传统出版模式以外的另一种选择。通过新的数字技术和网络化通信，任何人都可以及时、免费、不受任何限制地通过网络获取各类文献，包括经过同行评议过的期刊文章、参考文献、技术报告、学位论文等全文信息，用于科研教育及其他活动。从而促进科学信息的广泛传播，学术信息的交流与出版，提升科学研究的共利用程度，保障科学信息的长期保存。这是一种新的学术信息交流方法，作者提交作品不期望得到直接的金钱回报，而是提供这些作品使公众可以在公共网络上利用^[8]。

基于以上定义，对用户免费以及无限制的合理使用是开放获取核心所在。开放获取不是免费获取（free access），也不是普遍存取（universal access），而主要有以下三方面特征：①通过网络获取，强调同行评议。开放获取的期刊一定是同行评议的期刊，而不是自由获取（free access）。开放获取期刊一定是同行评议的，一定是免费获取的。但是免费的期刊不一定是同行评议的期刊，也不一定是开放获取期刊；②强调他人（作者、机构）付费出版，读

者免费使用，强调读者利用上没有经济、法律、技术的限制；读者可以获得全文，而不是文摘；③强调作者和版权所有者授予全世界所有读者免费、永久地获取、复制传播、向公众展示作品、传播派生作品、以合理的目的将作品复制到任何形式的数字媒介上的权利，以及用户制作少数印本作为个人使用的权利。开放获取的附加特征，包括自存储时的时滞、永久获取、版权所有者认同、作者拥有版权、遵循有关的开放获取的版权协议（如 Creative Commons、OAI 等）^[7]。开放获取的范畴，不仅包括期刊论文，还包括软件、科学数据、学位论文、教学课件、图书、音频、视频、多媒体等等。表 1.1 总结了几个重要 OA 计划中对 OA 定义的界定与说明。

表 1.1 OA 重要计划对 OA 定义的界定与说明

| 时间 | 来源 | 权利人 | 限定条件 | 权利 |
|------|--|-------------------|--------------------------------|---|
| 2002 | 布达佩斯 开放获取 计划 | 作者 | | 署名权、保持作品完整权 |
| | | 所有用户 | 合法用途 | 阅读、下载、复制、传播、打印、检索、链接、导入软件 |
| 2003 | 关于 OA 出版的 《百思达宣 言》 | 作者及著 作权持 有者 | 免费、不可撤销、 世界范围、永久的 | 授予使用者权利、恰当署名、以标准格式至少在一个由声誉卓著的机构支持的在线仓储中公开存储 |
| | | 所有用户 | 不限理由、任何数 字形式、个人使用 的少量印刷本 | 公开复制、使用、传播、展示原作品，并制作和传播衍生作品 |
| 2003 | 关于自然 与人文科 学知识开 放获取的 《柏林宣 言》 | 作者及著 作权持 有者 | 免费、不可撤销、 全球性权利 | 署名作者，将包括各种附件在内的完整作品至少提交到一个应用 OAI 标准的在线仓储中，该仓储必须由学术机构、学会、政府部门或者其他知名组织设立和管理，以确保长期可用 |
| | | 所有用户 | 不限目的、任何数 字形式、个人使用 的少量印刷本 | 公开复制、使用、分发、传播、展示原作品，传播衍生作品 |

续表

| 时间 | 来源 | 权利人 | 限定条件 | 权利 |
|------|---------------------------|-------------------------------------|----------------------------------|--|
| 2004 | IFLA/ Welcome Trust | 作者及著作 权持有者 | 免费、不可撤销、 世界范围、(著作权 保护期内)永久 | 署名作者、必须将完整的作品和附属 材料,加上上述内容的许可协议,以 标准电子格式,在作品出版后立即、 至少在一个应用互操作协议、不限制 传播、能够永久保存的学术机构、政 府部门、其他允许 OA 的组织仓储中 存储 |
| | | 使用者 | 以任何合理目的、 署名作者、为个人 使用少量复制 | 复制、使用、传播、表演、公开展 示、衍生任何载体的作品 |
| 2006 | 澳大利 亚 OAK Law | OA 的核心原则是开放获取研究和学术成果,尤其是公共基金支持的研究成果 | | |

1.1.3 开放获取发展

在开放获取目录网站 (Open Access Directory, OAD), 完整记录了开放获取发展历程中重要的时间表与事件^[9], 其中对开放获取发展影响重要的历史事件笔者介绍如下。

1998 年, “自由扩散科学成果运动” (也被称为“自由科学运动”) 是公认的开放获取标志性时间^[9], 倡议对于科学文献要减少版权条约中的限制条款, 反对将作品复制权从作者转移给出版商。该倡议被认为是开放获取的重要起点, 并逐渐引起全球广泛关注。

2001 年 12 月 1—2 日, 开放科学研究所 (Open Society Institute, OSI) 在匈牙利首都布达佩斯召开的科学和学术文献开放获取会议上发布了“布达佩斯开放获取计划”^[5], 提出推动科技文献的开放获取, 即用户通过互联网免费阅读、下载、复制和传播作品, 并提出开放获取的两种实现形式: 开放存储



(Self-Archiving) 和开放出版 (Open-access Journals)。开放存储即作者在论文发表后将论文存储在机构或专业知识库, 立即或者延迟一段时间 (一般为6个月或者12个月) 后开放发表; 开放出版即期刊以免费获取方式发表论文, 供公众开放阅读。布达佩斯会议督促国际社会加快促进各学科领域的研究论文在因特网上免费获取的进程, 从此引起国际社会的关注、支持、响应和参与。人们已经基本对开放获取的内涵、组织形式达成了共识, 此会议标志着开放获取进入快速发展阶段。

2003年10月, 德国马普学会发起柏林会议, 通过《关于自然科学与人文科学资源开放获取的柏林宣言》^[10], 德国、法国、意大利等多国的科研机构于2003年10月22日在柏林联合签署, 旨在利用互联网整合全球人类的科学与文化财产, 为来自各国的研究者与网络使用者在更广泛的领域内提供一个免费的、更加开放的科研环境; 呼吁向所有网络使用者免费开放更多的科学资源, 以促进更好地利用互联网进行科学交流与出版。将开放获取的对象扩展到科研论文、科学数据、参考资料、图片图表、学术类多媒体资源等, 并提出了开放获取的两个条件, 即作者或者版权所有人承诺向所有用户提供免费的、不能撤回的和在全世界范围内复制、利用、传播权利, 只要保证以适当方式显示作者权利; 作品的完整版本应该以标准格式存储到在线存储库中以支持作品的开放获取和长期保存。

随着科学数据、教育资源等广泛数字化, 开放获取得到进一步发展。2004年1月, 世界经合组织 (Organization for Economic Co-operation and Development, OECD) 科技政策委员会发布由34个国家联合签署的《公共资助研究数据的开放获取宣言》^[11], 提出公共资助项目产生的科研数据应实现开放获取。该宣言提出要建立公共资助研究数据的开放获取机制, 要建立透明的关于数据存储权、拥有权、使用伦理及其他限制、知识产权保护、使用者责任等的说明规则, 通过标准化提高数据的互操作性, 建立搜集和传播数据的

最佳方法和技术来确保数据的真实性、原始性、安全性和完整性，建立数据管理最佳方法和专门服务来促进全球数据共享的高效率等，同时要兼顾开放获取为科研创新带来的利益和保护合法权益所需要的限制间的平衡，在设计科研数据获取机制时应与国家的法律体系保持一致。

2006年12月，OECD又颁布了《公共资助研究数据的开放获取原则和指南》^[12]，指出“公共资助的研究数据”是由政府机构研究获取的数据、或利用任何级别的政府资金资助的研究产生的数据；科学数据主要指用来作为科学研究主要来源的事实记录（数值分数、文字记录、图像和声音等）。实际上，科学数据开放获取中已经涉及关于数据收集、处理、利用、保存、管理等一系列机制，已经成为科技基础设施建设的组成部分。

2007年9月，在南非开普敦举行的开放教育会议发布《开普敦开放教育宣言》^[13]提出，促进教育领域的开放资源、技术和教学实践。2008年，欧洲大学协会理事会批准了其开放获取工作组的建议，要求各国大学校长理事会积极建立支持开放获取的政策机制。2009年美国国家医学院发表《美国对世界健康的承诺：对新行政当局的建议》，提出公共资金资助的健康研究成果应该为公众开放获取。2009年，白宫科技政策办公室公开征询公众意见，将国立健康研究院实行的资助项目研究论文强制性开放存储政策推广到整个联邦政府。2010年4月，美国众议院六位议员提出《联邦政府研究成果公共获取法案》^[14]，要求所有资助金额达到一亿美元以上的联邦资助机构实施与国立健康研究院相同的开放存储政策。2010年5月，包括哈佛、斯坦福、普林斯顿、康奈尔、卡耐基-梅隆、加州大学及其各个分校、宾州、杜克等27所研究型大学教务长集体呼吁国会支持《联邦政府研究成果公共获取法案》。

2016年3月21日，德国马普学会等机构通过设立专门网站正式发起了“开放获取2020（Open Access 2020，OA 2020）”国际行动计划。OA 2020，是在《开放获取柏林宣言》框架下和基于开放获取使命声明的国际行动计划，