

机构知识库建设机制初探

洪梅 马建霞

(中国科学院国家科学图书馆兰州分馆, 730000)

【摘要】介绍了在开放获取的大环境下, 机构知识库的概念及意义; 重点探讨了机构知识库的机制建设中面临的问题和挑战, 用机制保证机构知识库的建设和运营成功。

【关键词】机构知识库 机制 开放获取

21 世纪是信息的时代。随着信息高速公路的建设, 信息共享作为人们的需要, 已成为一种趋势。开放获取作为一种新的、高效的交流模式, 发展的时间尽管较晚, 但势头强劲, 很好的推动了信息共享。

1 开放获取

1.1 开放获取的含义

根据 BOAI 的完整定义^[1], “提供不同程度和种类的更广阔更方便的获取文献资料的途径。对于文献资料的‘开放获取’, 是指可以免费在公共网上得到, 允许读者阅读、下载、复制、传播、打印、检索或者链接文章的全文, 可以制作索引, 把它们存入数据库, 或者为了其他合法目的的使用, 除了保持作品的完整性外, 不存在资金、法律、技术障碍。对复制和传播方面唯一的限制, 或者说版权的唯一作用是让作者能够控制其作品的完整性, 使其作品能够被承认和适当地被引用。”开放获取打破了两层壁垒: 第一, 免费在公共网上得到, 消除了价格壁垒; 第二, 允许很大程度的使用, 打破了使用权限壁垒。

1.2 开放获取的途径

开放获取的途径主要分为两大类:

(1) “金色通道”: 即开放获取期刊(OAJ)。现有期刊转为开放获取期刊, 或创办新开放获取期刊, 对作者收取出版费用, 而不是对用户收取使用费用, 免费提供学术期刊的在线版本。

据《开放获取期刊目录》(DOAJ)统计, 截止到 2006 年 11 月 10 日, DOAJ 所收录的开放获取期刊已达 2459 种, 有 119979 篇文献提供于 DOAJ 服务中^[2]。

(2) “绿色通道”: 即开放获取自存档(OA self-archiving)。发表过的文献资料由作者自己或第三方将作品存储在作者个人网站、学科知识库(disciplinary archive)或机构知识库(institutional repository)中, 以提供免费获取, 对开放获取起着杠杆调节的作用。

1.3 开放获取的政策

国际开放获取运动如火如荼的开展着, 各个国家机构组织纷纷起草签署宣言, 以更好地利用网络进行科学交流与共享。

根据布达佩斯开放存取计划(Budapest Open Access Initiative)的精神, 由德国马普学会发起《关于自然科学与人文科学资源的开放使用的柏林宣言》(Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities), 是最具有国际影响的宣言之一, 旨^[3]在利用网络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。中国科学院和国家自然科学基金委员会于 2004 年 5 月 24 日分别签署了该宣言^[4], 标志着我国正式参与到国际化的开放获取运动中去。目前^[5], 已有全世界 185 个组织签署了该宣言。美国的 NIH 实行的“公共获取政策”^[6], 要求受 NIH 资助的对象, 在正式出版 12 个月内必须把最后的同行评议过的版本提交到

PMC (PubMed Central) 知识库中。

2 机构知识库的含义

机构知识库的概念主要源于两个方面^[7]: (1) 基于“开放存档先导计划”(OAI)和其推出的“用于元数据收集的开放的文档存储协议”(OAI-PMH)的数字互操作性;(2) 类似与图书馆常规功能的收集、存储、分类、编目, 提供以数字对象为内容的获取的数字图书馆。

对于什么是机构知识库说法不一, 但究其本质, 机构知识库就是^[8]机构依托网络运作的一组资源管理与服务机制, 它以本机构成员在工作过程中创建的各种知识产品(如科学数据、科研文献、科研成果、教案、讲稿等资源), 尤其是以数字对象为内容, 以对这些对象的捕获、收集、整理、集成、描述、标识、组织、索引为途径, 以实现对这些对象的永久保存和广泛传播为目标。本质上, 机构知识库就是收集, 存储, 传播信息的工具。

机构知识库的目的^[9]主要有两个: 第一, 通过自存档提供开放获取文献输出, 增加文献的透明度, 具有开放和互操作的性质; 第二, 长期保存机构的数字资源, 包括未发表的和灰色文献, 如论文、技术报告等。

3 机构知识库的机制

建立机构知识库尽可能得实现了数字资源的长期保存, 有利于信息快速的传播, 降低了获取文献的成本, 提高了学术成就的影响力。从某种意义上说, 机构知识库的成果, 反映了开放获取运动的成果。但是在机构知识库的实际建设和运营中, 面临很多障碍和挑战。

3.1 项目组织

机构知识库的建设, 是一个较为庞大的工程, 需要有计划有目的的组织这个项目, 涉及到机构多方面的人员配合完成, 主要包括政策制定人员、信息技术人员、专业研究人员和图书馆员。

对于政策制定人员, 一方面, 应制定相应的激励机制, 激发作者提交其作品的积极性和主动性; 另一方面, 应制定相应的许可协议, 确保用户合理有效的利用作品而不侵犯其知识产权; 另外, 应制定相应的服务和标准, 使机构各方面人员有序有据的开展工作。

对于信息技术人员, 应为机构知识库搭建一个满意的服务平台, 以满足各个方面的需要。

对于专业研究人员, 应在图书馆员需要的时候, 提供他们的帮助。

对于图书馆员, 应协助作者, 特别是工作太繁忙或有技术障碍的作者提交作品, 对于已出版的符合机构需要的作品, 在版权许可的条件下, 进行数据的输入, 应保证作者提交元数据的正确性; 应及时向政策制定人员反映在机构知识库的运营过程的遇到的新问题, 以为制订新的政策提供依据; 应向信息技术人员提出需求, 以便在技术上改进; 应在专业知识模糊不清时, 向专业研究人员请教, 保证工作的精确性。另外, 图书馆员依附图书馆对外宣传推广机构知识库。

3.2 知识产权

版权作为知识产权的一部分, 在开放获取运动中, 是一个讨论的焦点。建立机构知识库, 版权的许可问题, 越来越受到人们关注。

在开放获取运动中, 开放的是作品的内容, 目的是为了知识的传播, 作品的版权作为一项基本权利是保留的。作品的版权可能属于机构, 可能属于作者本人, 也可能属于出版商。在作品发表前, 若作品的版权属于作者本人, 作者可自愿进行自存档; 若作品的版权属于机构, 机构可自行存档。在作品发表后, 作品的版权属于出版商, 在存档前, 须经过出版社的许可。尽管有证据证明许多期刊出版商是支持自存档的, 但是可能会对出版的版本进行限制,

如不允许提交最后出版商制作的 PDF 格式的版本,这就需要作者提供最后版本^[10]。在荷兰,一项名为“科学奶酪”^[11]的项目中,关于平衡作者和出版商的版权问题就有很好的规定:一方面,出版商对涉及文献商业行为的版权具有唯一许可权,另一方面,作者拥有涉及文献学术用途的许可权。不论是预印本,还是已出版发行的文献,都要保证作者的最后出版的版本至多 6 个月存档于知识库。有的机构知识库,在记录中有版权栏字段。如香港科技大学的机构知识库^[12],有字段 "[Journal title] © copyright (year) American Physical Society. The Journal's web site is located at http://.....",就充分的显示了该记录的合法性,表明该记录经过期刊出版社的许可的。

知识产权不仅仅存档的版权许可,还包括谁来使用这些存储的资料,如何使用等等。对于用户的最低要求,是要遵循 CC 许可协议^[13],机构还可以根据自身的需要制定协议。

在机构知识库中可开放获取作品,不少行为就需要通过一些技术手段来控制。如奇迹文库^[14],为了保证大多数用户的权益,就暂时屏蔽同时、批量下载多个文件的用户 IP 并提倡更多地采用 P2P 的下载方式。

3.3 运营费用

机构知识库要长期存在并发展壮大,必须要有资金作为基础。这些资金主要用于机构知识库的政策起草,许可协议的制定,指导方针的改进,用户培训,元数据的建立等等。由于机构知识库提供的是开放获取,每一个用户可得到机构知识库里的记录,而不收取额外的费用。机构知识库的运营所需要的费用就很自然的转嫁到机构和作者头上。机构在建立知识库的初期,就要有一定的财政预算,以保证知识库的正常运行。在机构知识库的运营过程中,可向政府、协会、组织寻求外援资助。此外,收取作者出版费用是目前开放获取期刊最主要的成本弥补手段,常被简称为“作者付费出版,读者免费使用”,动员作者从研究费用中支出一部分作为发表费用,也可以成为机构知识库运营费用的来源之一。

3.4 质量控制

机构鼓励作者自存档作品于机构知识库,其质量的控制,尤显得重要。作者提交作品到机构知识库,机构知识库只是对作品进行存档,只有评审通过了的作品,才被机构知识库所承认以提供开放获取。评审的内容包括作品是否符合该机构的服务对象的需求,作品是否属于所要存档的文献类型,作品是否有版权许可,作品是否符合一定的格式,作品的元数据是否符合一定标准等。

建立一个机构知识库,它的服务对象首先要明确。如果某大学建立的机构知识库,那么它的服务对象就是该大学的全校师生员工等校园成员。如果象英国那样建立国家级的机构知识库,那么它的服务对象就是全国的需要利用网络进行学术交流的人员。机构知识库存档的作品,主要是为特定的对象服务的。可以从作品所属的学科来对作品进行归档,以判断作品是否符合该机构的服务对象的需求。

作品的文献类型是作品的一个重要属性,可以较好的揭示作品的基本情况。一般需要存档的是论文预印本、已发表论文、技术报告、科研进展报告、硕博士论文、图书或其章节、科研数据库、会议录、教学资料或课件、计算机软件、其他灰色文献等等,而行政记录、与机构的历史和活动有关的一些资料等则可以存储在机构的档案馆^[15]。

只有拥有版权许可的作品,才可能通过评审,以保证作品的知识产权。

为了更好的长期保存,统一管理,机构协作,对于上传作品的内容描述规定必填项以生成元数据记录,对于上传作品的格式做统一的规定,如图书馆情报学开放文库的规定^[16],对于上传文献的格式统一规定为“pdf”,因为 PDF 可以不受语言和计算机环境的限制,保证不同地区的人们都能够看到文件内容,而不是乱码。在 Eprints 的示范站点^[17]中,规定了

四种可提交的格式：HTML、PDF、Postscript、ASCII。

作品的元数据直接揭示作品的内容，在作者提交作品之后，需要由图书馆员保证元数据的正确性，对于错误的元数据，图书馆员有权进行修改或拒绝其提交。

此外，机构知识库还应对通过评审存储到机构知识库提供开放获取的作品建立反馈机制，通过用户评议，及时删除更新，以保证机构知识库的整体质量，使机构知识库更好的发挥其作用，而不是使用户迷失在信息的海洋里。

3.5 激励机制

一个机构知识库要成功的运营下去，制定一些激励机制的项目和政策是十分有必要的。在国外^[18]，围绕电子版论文和学位论文所制定的项目与政策，已经促进建设了许多机构知识库。比利时采取的一项激励措施，是为作者提供建立和维护个人出版物目录的服务；而在荷兰，则为作者所存入的成果提供一个到皇家图书馆 E-Depot 的链接，在项目“科学奶酪”^[19]中建立的国家级的机构知识库，里面尽可能多的收录了数百位国家学术带头人的公开发表的成果，为机构知识库带来了声望和地位，很好的激励作者提交他们的作品。

在国内，也应把如何激励作者自愿提交他们的作品作为一个战略目标。有的作者因为在技术操作上有困难，或难以忍受提交过程的等待；有的作者很难接受开放获取，也不觉得自己能从机构知识库中得到什么好处，所以对于提交自己的作品，呈现出消极甚至是反对的态度。对于这些作者，一方面，通过技术方面的支持，尽力减少科研人员和教师在提交资料时的工作量；更重要的是需要机构进行有针对性的培训，从技术和心理方面消除他们的障碍，使他们能够方便快捷地使用机构知识库，看到机构知识库能够提供的服务，如引用分析，阅读情况，个人发表记录等，以实现更好的传播和利用作品；如果有可能的话，制定与项目管理和科研管理相结合的政策，在研究室和研究项目的考核中，制定必须向机构知识库提交相关的知识产出的政策，以确保研究人员主动配合机构知识库的资源提交。

有人提议，在宣传过程中，我们应该推广的是“机构知识存储和服务”平台，而不是 Dspace 或是 E-Prints 等，因为后者只是一个基础技术平台。我们应该将应用程序和服务名称相区分开，应用程序可以更换，服务名称却可以相对固定，这样有利于提高服务品牌的用户忠诚度。

另外，向 Google 和类似的搜索引擎开放机构知识库的元数据，使得作者所提交的作品更容易被查到和获取。通过这些手段，既可证明阅读率和引用率，也可以帮助营造作者自存档的氛围。

3.6 工作流程

作者在注册成功后，就可提交作品到机构知识库。经过评审，符合各方面要求的作品存储到机构知识库，若有不符合某方面要求的作品，图书馆员按其权限予以修改至符合要求，若其不符合的方面超出图书馆员的修改权限，则通知作者，将作品退回。图书馆员根据反馈机制，按一定的标准，对机构知识库里的作品进行整理。任何用户都可以访问到作品元数据，对访问全文的用户，根据实际情况划分不同的等级加以控制，但是如需服务，如 E-mail 通知更新情况，就需要注册。

对于已出版的作品，若具备了相应的许可，图书馆员可批量导入到机构知识库中。

4 结语

机构知识库作为开放获取运动的一个重要部分，想要发展壮大，决不仅仅是开发一个软件，把内容往里面放这么简单。它需要人力、物力、财力等各个方面的配合，机制的健全就尤显得重要。国外和中国香港已经有了很多例子值得我们学习，他们所面临的障碍和挑战，可能也是我国的机构知识库的已经或者将要面临的障碍和挑战。可喜的是，我国大陆已经有

一些大学和研究性图书馆开始建立自己的机构知识库。图书馆作为知识传承、学术传播、信息服务的主体，将在运营机构知识库和建立机构知识库的机制方面起着引领作用。

参考文献:

- 1 [2006-10-27].<http://www.soros.org/openaccess/read.shtml>
- 2 [2006-11-10].<http://www.doaj.org>
- 3 [2006-12-4].<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>
- 4 [2006-12-4].<http://www.cas.ac.cn/html/Dir/2004/05/25/6730.htm>
- 5 [2006-12-4].<http://oa.mpg.de/openaccess-berlin/signatories.html#top>
- 6 [2006-12-5].<http://publicaccess.nih.gov/index.htm>
- 7 9 [2006-11-7].http://en.wikipedia.org/wiki/Institutional_repository
- 8 常唯. 数字学术环境中的机构知识库探析. 图书情报工作[J], 2006, 50(7): 46-50
- 10 A Librarian's guide to institutional repositories. [2006-11-14]. <https://magpie.lboro.ac.uk/dspace/bitstream/2134/1122/1/Pickton&Barwick+Elucidate%5B1%5D.pdf>
- 11 [2006-11-18].<http://www.surf.nl/copyrighttoolbox/authors/licence/>
- 12 陈丽霞. 建立和维护香港科技大学的机构学术成果存储库. [2006-11-14]. <http://repository.ust.hk/dspace/handle/1783.1/2425>
- 13 [2006-11-20].<http://creativecommons.org/licenses/by-nc-sa/1.0/deed.zh>
- 14 [2006-11-5].<http://www.qiji.cn/scinews/detailed/127.html>
- 15 Raym Crow . The Case for Institutional Repositories: A SPARC Position Paper . [2006-11-23].<http://www.arl.org/sparc/IR/ir.html>
- 16 [2006-11-5].<http://dspace.nlc.gov.cn/>
- 17 [2006-12-3].http://demoprints.eprints.org/perl/users/submit?eprintid=7596&_action_submit=1
- 18 Gerard van Westrienen, Clifford A. Lynch. Academic Institutional Repositories-Deployment Status in 13 Nations as of Mid 2005. [2006-11-16].<http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>
- 19 [2006-11-18].<http://www.creamofscience.org>

作者简介:

洪梅, 女, 1983年生, 中国科学院国家科学图书馆兰州分馆2005级情报学硕士研究生。
马建霞, 女, 1972年生, 毕业于中科院研究生院·中科院文献情报中心, 图书馆学, 管理学硕士。工作单位: 中国科学院国家科学图书馆兰州分馆, 副研究馆员, 硕士生导师。发表论文二十余篇。研究方向: 数字图书馆理论与实践、专题信息服务系统研究与开发。 email: majx@lzb.ac.cn tel: 0931-8272239ⁱ

ⁱ基金项目: 本文受中国科学院西部之光人才培养计划暨甘肃省自然科学基金项目3YS051-A25-001资助。