

信息开放获取新平台——机构知识库

徐以鸿 朱 涛

(中国科学院力学研究所图书信息中心,北京,100190)

摘 要 概述了机构知识库的定义,阐明了图书馆建设机构知识库的重要意义,简要介绍了中国科学院力学研究所图书馆机构知识库的建设情况。

关键词 机构知识库;知识共享;开放获取;图书馆

中图分类号 G250.74 **文献标识码** A

为了促进科研成果在互联网上的自由传播,在国际学术界、出版界和图书情报界的推动下,20世纪90年代,开放存取运动开始兴起。随着互联网的普及和电子出版技术的完善,开放存取运动在这短短的十几年取得了长足的发展。在开放存取运动的发展过程中,主要形成了3个分支,即开放存取期刊、学科知识库和机构知识库。在这3个分支中,机构知识库的发展最为迅速。中国科学院下的各研究所是国内学术资源的重要产出地,科研人员和研究生发表的会议论文、期刊论文、学位论文、论文预印本、研究报告、技术报告项目中的申请报告、单位的新闻报道、会议纪要和档案、图片文件、声像文件、数据、课件、学习对象等都是重要的学术资源,是机构的知识资产。但这些知识资产很多都散落在机构各部门、实验室或课题组及个人手中,其他部门和机构获取使用很困难。如果由机构集中保存和提供使用,则可使全单位的科研人员甚至全社会获益。

1 机构知识库的定义

机构知识库也称机构资料库、机构库或机构典藏。机构知识库(Institutional Repositories,以下简称IR)是利用网络及相关技术收集、整理和保存一个机构或多个机构所生产的知识产品,形成一个数字形式的学术数据库,允许机构及机构外的成员通过互联网免费或授权获取使用。IR中的资源种类不限,如期刊论文、会议论文、学位论文、技术报告、实验报告、手稿、胶片、讲义、课件等,资源形式也不限,可以是文档类、图像类以及超链接等。

2 建设机构知识库的意义

2.1 为研究机构提供一种新的资源整合方式

中国科学院力学研究所图书馆经过近几年的快速发展,已建立起纸本资源、电子资源和馆际互借及文献传递服务支撑的三级文献保障体系,但对本研究机构最具特色的资源,即学术产出成果如何保存、传播、使用还没有一个有效的方法。机构知识库的建立就可把研究人员的学术论文、研究报告、实验报告、专利、学位论文等整合到机构知识库中,由图书馆进行统一管理,为研究所提供了一种新的资源整合方式。

2.2 提升机构的学术影响力

通过机构知识库把研究所的所有科研成果、科研活动以及机构的公共服务都可以展示出来,并能从机构整体的角度反映机构的智力资源。机构知识库的实施可以确保研究成果从整体上被广泛传播和长期保存、提供利用。在网络数字化技术日益成熟,纸本与电子学术期刊费用不断提高的情况下,通过建立IR,各个学术机构能将研究者的成果进行集中保存,并开放给其他

学者参考使用,有助于提高机构在学术领域的知名度和显示度。

2.3 促进知识共享和知识更新

机构知识库为研究所提供了一种新的有效的资源整合方式,其包含的资源可向机构内外的成员提供开放获取服务。机构知识库的建设和实施可以避免传统出版方式所带来的出版时滞,实现学术领域内科技信息与研究成果的交流,促进科研的创新和知识的共享。

3 机构知识库实施前的准备工作

3.1 制定相应的政策和规则

在建立机构知识库之前,必须详细分析机构所拥有的专业知识资源和所需的知识资源,明确机构知识库内资源的类型,以便清晰地划分整个机构知识库的内容和结构,进而合理地规划建设机构知识库。

机构知识库一般设在机构局域网内,应由两部分组成:一是可公开获取及传播的显性资源,如图书、期刊论文、会议论文、学位论文、科技报告、专利以及标准等;二是只在机构内部传播的隐性知识资源,如科研人员的课件、手稿、实验数据、计算程序、设计图纸、学习资料、会议纪要、新闻、图片、图表以及音频、视频数据等。

3.2 构建机构知识库管理平台

目前,用于机构知识库创建的工具比较多,有Fedora(由Univ. of Virginia and Cornell开发)、Eprint(由University of Southampton开发)、Dspace(由MIT和HP联合开发)以及GreenStone(由University of Waikato, Witten & Bainbridge开发)等。以上这些工具有优缺点,我们决定采用由MIT和HP联合开发的Dspace作为构建机构知识库平台是基于以下几点考虑:一是Dspace是全免费的基于BSD open source license的一个开放性源代码的软件平台,遵循Mozilla协议,可自由使用、复制、修改,机构通过直接利用或改写软件方式,便可利用Dspace创建自己的机构知识库;二是Dspace使用的第三方软件也是开源代码,如PostgreSQL, JDK, Apache等;三是Dspace具有良好的功能和性能,使用高质量的都柏林核心元数据提供受控词汇,并且使用了URLs,这是一种持久稳固的网络标志符;四是Dspace目前仍然有计划地、持续地推出功能不断完善的新版本,且Dspace设有官方网站和论坛,使用人员可以与官方相互沟通。

3.3 准备所需的硬件和外围设备

机构知识库作为机构永久保存知识资源的仓库以及知识共享的平台,必须考虑数据的稳定性和永久性,所以需要一台独立的服务器,以保证工作人员容易对其升级更新或进行实验而不

中断其他服务。另外,可以通过 RAID 冗余容错技术解决数据的备份问题,同时还要考虑工作人员终端计算机的处理速度和网络的带宽及速度。

3.4 图书馆在机构知识库建设中的作用

图书馆是研究机构的信息中心、知识中心,在信息加工和管理方面具有独特的优势,尤其是现今很多科研单位图书馆与信息网络部门合并在一起组成图书信息中心,这就使得图书馆在构建机构知识库时能够起到重要的作用。另外,图书馆还可对机构知识库的内容进行长期保存,所以图书馆在机构知识库的建设管理中理应承担主角。

4 中国科学院力学研究所图书馆机构知识库的建设

近年来,随着开放获取技术的长足发展,机构知识库的建设也在国内外迅速展开,据统计已建立了机构知识库的国家有澳大利亚、加拿大、美国、比利时、法国、英国、丹麦、挪威、瑞典、芬兰、德国、意大利、荷兰以及我国的香港地区,机构知识库建设比例较高的国家有德国、挪威、荷兰和澳大利亚。在这 4 个国家中,除了澳大利亚有 5% 的大学还没有建立机构知识库外,其他 3 个国家所有的大学都已经建设了自己的机构知识库。这表明这些国家的大学非常重视机构知识库的建设,而且也投入了很多的人力、财力来支持这项工作。对于这些国家来说,目前已经可以开展其他相关方面的基础建设,以使其机构知识库达到全面实用的水平。国内目前已建设机构知识库或类似系统的机构有香港大学、香港科技大学、清华大学、中国科学院国家科学图书馆兰州分馆、中国农业大学等。

中国科学院力学研究所图书馆的机构知识库是经多方努力才建立起来的。它以 MIT 和 HP 联合开发的 Dspace 软件作为平台,以 Tomcat 作为 Web Server,以 PostgreSQL 作为 Dspace 资料库。机构知识库是围绕研究社群来进行组织的,每个研究社群可以分为多个子社群和专题,其数量不受限制。每个专题下则可以包含若干条目(Items),每个条目都有一个元数据描述。

目前力学所的机构知识库由专著、期刊论文、会议论文和学位论文 4 个专题组成,提供了匿名用户访问、授权用户提交内容、用户审核和管理等功能。

(1)检索及浏览:机构知识库提供了基本检索和高级检索。基本检索是输入一个检索词后,系统就会在题名、关键词、作者、摘要以及全文文档中进行检索。高级检索提供使用布尔逻辑来精确地查找资料,可以按研究社群/专题、题名、作者等浏览。

(2)用户提交内容:用户要想提交资料到机构知识库,首先必须注册成为许可用户,然后输入用户的 E-mail 和密码登录到 IR 系统,按提交的流程一步步操作,完成内容的提交,待管理者审核后就可浏览了。

(3)管理员功能:管理员可以新建、编辑、删除研究社群/专题和给专题授权,新建用户和用户组,并给用户和用户组授权,导入用户和数据,编辑新闻和数据,审核数据和删除数据。

力学所机构知识库的建设还处于试验阶段,所有的数据都是图书馆工作人员从其他数据库中导入的。今后的工作重点是一方面进行平台的完善建设,另一方面对数据加以补充,增加更多的专题,让科研人员真正加入到机构知识库的建设中来,一同把机构知识库建好、维护好。

5 结语

虽然机构知识库建设取得了迅速发展,但也应当看到,国内对机构知识库的建设与研究并没有引起足够的重视。从个人层面看,科研人员对机构知识库了解甚少,缺少对机构知识库参与的激励机制;从机构层面看,科研机构对实施机构知识库的价值及用途还不太认同,缺乏相应的发展策略。另外,就是建立机构知识库的技术问题以及版权许可、同行评议和质量控制等问题,但最大的障碍是大众的不信任,这直接影响了机构知识库资源来源的稳定性和质量的可靠性。机构知识库管理者要利用各种场合,宣传机构知识库的开放获取,以引起大家的关注和讨论,推动机构内所有的部门和人员积极参与机构知识库的资源创造、储存、共享和交流,以有效地实现机构知识资源的沉淀、积累和创造。从学术交流的全过程考虑,机构知识库事业的发展需要科研人员 and 科研机构、学术团体、图书馆及政府机构的共同努力,将机构知识库建设成图书馆的特色资源,成为研究所新的开放存取平台。

参考文献

- [1] 常唯.机构知识库:数字科研时代一种新的学术交流与共享方式[J].图书馆杂志,2005,24(3):16-19.
- [2] Emily Dill, Kristi L Palmer. IDeA: 实施机构知识库的思考[J]. 图书情报工作动态,2005(10):16-20.
- [3] 李麟.推动 21 世纪的学术信息交流:机构知识库在开放获取运动中的作用——2007 年 OSI/eIFL 开放获取会议综述[J]. 图书情报工作动态,2007(10):6-15.
- [4] 柯平,王颖洁.机构知识库的发展研究[J].图书馆论坛,2006,26(6):243-248.
- [5] 于佳亮,吴新年.机构知识库的研究及发展现状[J].现代情报,2007(5):49-52.

(责任编辑 邱娅男)

第一作者简介:徐以鸿,男,1969 年生,1992 年毕业于中国科技大学,助理研究员,中国科学院力学研究所图书信息中心,北京市北四环 15 号,100190.

The New Platform for the Open Access of Information ——Institutional Repository (IR)

XU Yi-hong, ZHU Tao

ABSTRACT: This paper describes the definition of IR, expounds the important significance of constructing IR by university library and briefly introduces the progress of the on-going construction of IR of the library of the Institute of Mechanics of Chinese Academy of Sciences.

KEY WORDS: IR; knowledge sharing; Open Access; library