

云南省教育厅科学研究基金项目（2012Y416）

机构知识库的 理论与实践

■ 郎庆华 著

 云南大学出版社
YUNNAN UNIVERSITY PRESS

云南省教育厅科学研究基金项目

机构知识库的 理论与实践

■ 郎庆华 著

 云南大学出版社
YUNNAN UNIVERSITY PRESS

图书在版编目 (CIP) 数据

机构知识库的理论与实践 / 郎庆华著. -- 昆明:

云南大学出版社, 2014

ISBN 978 - 7 - 5482 - 2000 - 8

I. ①机… II. ①郎… III. ①学术机构—知识库—研究
IV. ①G211 ②TP311.13

中国版本图书馆 CIP 数据核字 (2014) 第 105718 号



机构知识库的理论与实践

郎庆华 著

责任编辑: 李 红

封面设计: 周 畅

出版发行: 云南大学出版社

印 装: 昆明市五华区教育委员会印刷厂

开 本: 850mm × 1168mm 1/32

印 张: 11.125

字 数: 289 千

版 次: 2014 年 6 月第 1 版

印 次: 2014 年 6 月第 1 次印刷

书 号: ISBN 978 - 7 - 5482 - 2000 - 8

定 价: 42.00 元

社 址: 云南省昆明市翠湖北路 2 号云南大学英华园内

邮 编: 650091

发行电话: 0871 - 65033244 65031071

网 址: <http://www.ynu.edu.cn>

E - mail: market@ynu.edu.cn

序

机构知识库 (Institutional Repository, IR) 的产生与开放存取 (Open Access, OA) 息息相关。

多年来,传统学术期刊一直占据着学术交流的核心地位。然而,传统学术期刊价格的不断上涨,致使出版商与图书馆、作者、用户之间的矛盾日益突出,学术交流体系也面临着重整的局面。为了消除学术期刊价格危机和信息获取障碍,受西方人权观的影响和网络技术及其相关技术发展的推动,开放存取应运而生。而急剧增长的原生数字资源长期保存与利用,以及对隐性知识进行管理等的迫切要求,促使机构知识库这种崭新的学术交流模式破茧而出,并成为国际科技界、图书情报界、出版界、学术界等领域的研究重点。

目前,国内外掀起了机构知识库的建设浪潮,许多发达国家都纷纷把机构知识库提高到战略高度,并作为国家基础设施来进行建设,足可见国外对机构知识库的高度重视。国内由于开放存取起步的时间还不长,许多人开放存取的意识比较淡漠,因而对机构知识库的认识还不够深入。可喜的是,随着近年来中国科学院机构知识库网格、厦门大学学术典藏库和 CALIS 机构知识库等的逐步成熟及其大力的宣传、推广,机构知识库正逐渐被学界所认识,机构知识库的建设正成为数字图书馆发展的新趋势。

机构知识库的建设是一个系统工程,它涉及内容收集、政策建设、长期保存、知识产权、技术标准、质量控制、服务等方面,而机构知识库建设迫切需要系统的理论来进行指导。读了郎庆华老师

的《机构知识库的理论与实践》，我认为该书具有以下三个特点：

一是选题前沿。作为开放存取的两大实施策略之一，机构知识库是一个理论性和实践性比较强的研究课题，也是该学科领域的研究热点。从2005年，作者就长期关注开放存取研究领域的发展，并把机构知识库作为自己的主攻方向。这对于西部地方高校的作者来说是需要一定的勇气的，并且没有深厚的功底也是难以做到的。

二是系统性较强。机构知识库是一个系统工程，它涉及政策、技术、标准、知识产权、长期保存等方面，它是多种因素相互作用和共同推进的结果。作者没有孤立地研究机构知识库，而是对机构知识库所涉及的各个方面进行了系统的研究，加之作者长期收集了大量的资料，近年来又发表了一系列论文，这为本书的系统性写作奠定了良好的基础。

三是具有一定的创新性。机构知识库的理论与实践在探索之中。作者借鉴国内外的研究成果，对当前机构知识库建设面临的一些问题进行了探索，如内容收集、政策体系、长期保存、服务、评价与质量控制等，并对机构知识库的发展前景进行了展望，尤其是把机构知识库置于泛在知识环境下来进行探讨，表现出了作者研究的独到之处和宏观视野。

当然，如果作者能对机构知识库评价、服务模式再做进一步的探讨，那么本书将会显得更完整。尽管这样，本书对机构知识库的理论与实践进行了系统的总结、研究和探索，为我国机构知识库的建设提供了一定的借鉴，具有一定的学术价值。

在《机构知识库的理论与实践》即将付梓之际，衷心祝愿作者在今后的科研道路上继续努力，不断取得新的成绩。

张绍武

2014年3月1日于曲靖

前 言

机构知识库 (Institutional Repository, IR) 是开放存取 (Open Access, OA) 运动的两大实施策略之一, 也是开放存取运动的产物。

在网络环境下, 随着机构原生数字化资源的急剧增长, 以及传统期刊价格的大幅度上涨, 催生了开放存取的理念。尤其是 20 世纪 90 年代, 由美国兴起的开放存取运动席卷全球, 掀起了学术交流模式的新变革。1991 年, 美国 LOS Alamos 国家实验室的物理学家 Paul Ginsparg 启动了机构知识库的雏形 arXiv.org。arXiv.org 促使物理学、计算机科学等相关学科领域的科研论文可以通过网络免费获取。此后, 开放存取理念逐渐深入人心。

为了推动开放存取运动的发展, 2001 年《布达佩斯开放存取先导计划》(Budapest Open Access Initiative, BOAI) 提出了开放存取的两种实现模式: 开放存取期刊和开放存取知识库。开放存取期刊被称为“金色之路”; 开放存取知识库被称为“绿色之路”, 其实现途径主要有学科知识库和机构知识库。其中, 开放存取期刊主要是用于解决数字学术资源的利用问题, 而开放存取知识库主要是用于解决数字学术资源的存储问题, 以便公共资金资助的研究成果能被公众免费获取。目前, 许多国家都纷纷掀起了机构知识库的建设高潮。这是因为机构知识库作为网络环境下的一种新型学术传播模式, 将对实施机构、图书馆、个人、学术交流、数字科研、社会、国家等都具有十分重要的意义。

在此背景下,笔者开始把开放存取中的机构知识库作为自己研究的主攻方向。2008年,随着自己第一篇有关机构知识库的论文被《情报杂志》录用,笔者开始对机构知识库的研究产生了浓厚的兴趣。由于当时机构知识库在国内的建设刚刚起步,因而还面临着质量控制、内容收集、政策建设、长期保存、知识产权、技术标准、服务等带来的诸多问题,而这些问题亟待我们去研究。为此,笔者于2009年和2012年分别申请并获准了两个云南省教育厅科研基金项目立项(“高校图书馆机构知识库的建设研究”,项目编号为09C0195;“高校图书馆机构知识库创新服务模式研究”,项目编号为2012Y416)。在这两个项目基金的资助下,笔者展开了对机构知识库的相关研究,并发表了一系列关于机构知识库方面的研究论文,加之多年来收集了大量相关的研究资料,使笔者对完成本书的写作有了充分的准备。几经寒暑,终于完成了此书。本书是“高校图书馆机构知识库创新服务模式研究”项目的研究成果之一。

目前,国内公开出版的有关机构知识库的学术专著有2009年李大玲的《学术机构知识库构建模式研究》、2012年李武的《开放存取的两种实现途径——OA期刊和OA知识库》和2013年谷秀洁的《开放型机构知识库著作权管理研究》三部。这些著作对机构知识库的某一方面进行了专门的、深入的研究,但尚未对机构知识库进行系统的研究。本书力图对机构知识库进行系统研究,并使之系统性与新颖性相统一,理论性与实践性相结合。

由于机构知识库是图书馆学情报学领域较为前沿的研究课题,囿于笔者学识和能力,本书难免会有一些纰漏和不足之处,因此恳请各位专家、学者和读者不吝赐教,以帮助笔者进步。

目 录

第一章 机构知识库概述	(1)
第一节 机构知识库产生的背景	(1)
第二节 机构知识库的定义、特征、功能、类型及其与其他学术交流模式的区别	(5)
第三节 机构知识库建设的意义	(18)
第四节 机构知识库的发展趋势	(26)
第二章 机构知识库建设的研究现状	(29)
第一节 国外机构知识库建设的研究现状	(29)
第二节 国内机构知识库建设的研究现状	(46)
第三章 机构知识库建设的实践与发展	(68)
第一节 国外机构知识库建设的实践与发展	(68)
第二节 国内机构知识库建设的实践与发展	(82)
第四章 机构知识库的构建	(101)
第一节 机构知识库的体系架构	(101)
第二节 机构知识库的构建步骤	(103)
第三节 机构知识库的构建模式	(113)

第四节 机构知识库构建软件的选择、评估与开发	(119)
第五节 机构知识库构建技术及其标准	(133)
第五章 机构知识库政策体系与版权保护	(145)
第一节 机构知识库的政策体系及其构建	(145)
第二节 机构知识库的版权保护	(160)
第六章 机构知识库的内容建设	(181)
第一节 机构知识库内容收集的类型及模式	(181)
第二节 机构知识库内容收集的现状及其影响因素 分析	(196)
第三节 机构知识库内容的收集策略	(200)
第七章 机构知识库的长期保存研究	(207)
第一节 机构知识库长期保存的价值和意义	(207)
第二节 机构知识库长期保存的研究现状	(210)
第三节 机构知识库长期保存的影响因素及其策略	(215)
第四节 机构知识库长期保存的发展趋势	(226)
第八章 机构知识库的服务研究	(235)
第一节 机构知识库服务的现状、问题及发展策略	(235)
第二节 机构知识库的服务模式	(243)

第三节	机构知识库服务的知识组织	(258)
第四节	基于知识管理的机构知识库服务体系的构建	(261)
第九章	机构知识库的评价与质量控制	(271)
第一节	机构知识库的评价	(271)
第二节	机构知识库内容的质量控制	(282)
第十章	机构知识库的发展前景展望	(295)
第一节	新一代机构知识库的特征与发展目标	(295)
第二节	泛在知识环境下机构知识库服务的发展策略	(299)
参考文献		(313)
后 记		(343)

第一章 机构知识库概述

机构知识库是开放存取运动的两大实施策略之一，是开放存取（Open Access, OA）运动实施的产物。

在网络环境下，随着机构原生数字化资源的急剧增长，以及传统期刊价格的大幅度上涨，催生了开放存取的理念。尤其是20世纪90年代，由美国兴起的开放存取运动席卷全球，掀起了学术交流模式的新变革。1991年，美国LOS Alamos国家实验室的物理学家Paul Ginsparg启动了arXiv.org。arXiv.org促使物理学、计算机科学等相关学科领域的科研论文可以通过网络免费获取。此后，开放存取理念逐渐深入人心。在此基础上，为了使机构公共资金资助的研究成果能被公众免费获取，以学术机构为轴心、基于开放存取理念的机构知识库便应运而生。

第一节 机构知识库产生的背景

为了推动开放存取运动的发展，《布达佩斯开放存取先导计划》（Budapest Open Access Initiative, BOAI）提出了开放存取的两种实现模式：开放存取期刊和开放存取知识库。开放存取期刊被称为“金色之路”；开放存取知识库被称为“绿色之路”，其实现途径主要有学科知识库和机构知识库。其中，开放存取期刊主要是用于解决数字学术资源的利用，而开放存取知识库主要是用于解决数字学术资源的存储问题。开放存取运动的兴起促进了

机构知识库的发展。机构知识库的产生并非一蹴而就，其产生背景可以追溯到以下几个方面。

一、网络信息技术的发展

网络信息技术的发展为机构知识库的产生提供了技术保障。

（一）互联网的普及与电子出版技术的完善^①

互联网的发展与应用无疑方便了学术界对信息的获取与利用。通过互联网，学者足不出户就可以进行信息的交流与传播，他们可以在互联网上随时发布信息，也可以通过互联网随时检索和浏览自己感兴趣的相关信息。同时，日益完善的电子出版技术也大大方便了学者对个人网站的建设以及自存档（self-archive）的实践，而学者的这些实践使得电子存储的想法在过去几年内得到学术界的广泛认同，并对建立机构知识库这种新的学术交流机制有着强烈的需求。可以说互联网的普及为机构知识库的建立提供了良好的网络基础，而电子出版技术的完善则为机构知识库提供了实践条件。

（二）数字资源管理技术

1. 网格技术和开放源代码的推动

网格计算源于元计算（Metacomputing），其初衷是将分布的多台超级计算机连接成为一个可远程控制和访问的元计算系统。网格体系根据高速和共享两大特征，正在不断向基于更大的网络、共享更多的资源、构建更强大的计算能力的目标迈进。网格技术将有利于解决大量访问、海量信息检索以及资源共享等方面的技术问题。而开放源代码的意义则在于将人类大量宝贵的创造力从“重复创造”中解放出来。近年来，开放源代码在数字图

^① 柯平、王颖洁：《机构知识库——大学图书馆的平台》，载《新世纪图书馆》2007年第1期，第1~8页、第99页。

书馆领域的应用已经相当普遍。如美国的 OCLC、英国的 UKOLN, 以及新西兰的 DL 项目都是图情领域推动开源代码应用的代表。网格技术和开放源代码为机构知识库的建立奠定了关键的技术基础, 使机构知识库思想的实现和快速发展成为可能。

2. 互操作协议和标准的出现

虽然电子出版技术的完善和互联网的普及使得机构知识库得以迅速发展, 但早期机构知识库存储的学术资源和其他网络资源一样, 存在格式不统一、位置分散等问题, 致使学术信息难以共享。为此, 学术界提出了 OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) 协议。该协议基于元数据收割的理念, 提供了一种独立于应用程序的互操作框架, 以便用户能在不同的机构知识库之间进行互操作, 以实现机构知识库间资源的共享。同时, OAIS (the Open Archival Information System) 参考模型、METS (the Metadata Encoding and Transmission Standard) 等标准与 OAI-PMH 协议一起, 也为机构知识库之间的互操作提供了标准, 方便了不同机构知识库之间资源的交叉检索和共享。

3. 基于数字对象的分布式信息系统架构的研究和实现

数字对象的研究和实现则为机构知识库的实施构造了理论框架。数字对象及基于数字对象的分布式信息系统模型最早是由美国人 Robert Kahn 和 Robert Wilensky 于 1995 年提出的。为了解决联邦数字图书馆的互操作问题, 他们提出了一种数字图书馆的基础建构块——数字对象, 通过数字对象可集成化地描述一个信息对象各方面的信息 (包括元数据), 并提供在网络环境下的统一存取。另外, 他们还提出了机构知识库的许多基本概念, 包括: 数字对象的唯一标识符 (handle)、数字对象的“类型”、仓储的概念、仓储访问协议 RAP (Repository Access Protocol)、数字对象服务请求传递和权限管理等。在后继研究者 Williams Arms、Carl Lagoze 等人的努力下, 这个理论框架逐渐成熟, 并最终形成

了数字仓储系统 Fedora, 这为机构知识库从理论变成现实提供了实施框架。

二、原生数字资源的长期保存与利用需求

自 20 世纪以来, 随着计算机技术、网络技术和数字化技术的迅速发展, 计算机和网络在教学与科研活动中的日益普及和应用, 使得大学和研究机构原来的学术环境日益数字化、网络化。科研人员和教师开始使用新技术进行学术研究活动、学术交流活动和教学活动, 从而使原生数字化资料 (born-digital material) 急剧增长。虽然这些原生资料对于科研活动有着重要的作用, 但这些原生资料中的绝大部分是不能通过传统出版方式出版的, 这使得原生资料的搜集保管工作难以进行。一方面, 机构个人或科研小组没有时间和技术来保存和传播这些数字资源; 另一方面, 机构成员的研究成果大多是以个人博客、个人网站、E-mail 等方式保存, 并随时处于“消失”状态。因此, 如何长期保存、有效管理与利用这些数字资产成为机构与学者面临的严峻问题。在与教师的交谈中, 麻省理工学院图书馆馆长 Wolpert 女士敏锐地意识到了建立机构知识库的必要性, 于是便促成了 DSpace 项目的产生。2002 年, 麻省理工大学 MIT 和惠普公司合作推出了 DSpace, 这表明了机构知识库由此诞生。

三、知识管理的重要性

20 世纪 90 年代, 随着知识的剧增以及知识组织和服务的需求而兴起知识管理运动。1998 年, Peter Drucker 在《哈佛商业评论》(*Harvard Business Review*) 上发表了一篇非常具有影响力的文章——《即将到来的组织》。他在文中指出, 知识已经成为现代企业最重要的资产, 只有那些能很好地管理和利用其知识资源的企业才能在市场竞争中占据优势。此文的发表引起了管理、管

理信息系统、信息管理等领域的学者对知识管理的研究。知识管理运动以及大量关于知识管理著作的出现,使得更多企业和机构的高层管理人员认识到信息资源的重要性,并积极投资知识库项目的研究与建设,这在很大程度上促进了机构知识库的产生与发展。

目前,很多组织都有专业化的知识管理工具,但这些知识管理工具主要是对显性知识进行管理。要达到知识管理的效果,还必须对隐性知识进行管理。隐性知识是由思维模式、信念和视觉构成的认知维度,难以清晰地表达,也很难用简单的工具进行管理。为了对隐性知识进行管理,激发组织内部创造、交换、使用知识的积极性和主动性,并进行知识共享工具、组织学习工具的相互配合,通过知识管理来实现知识创造,人们开始利用一些新的工具来管理信息和知识。而机构知识库的开源软件以其成本低、使用方便等特点,正受到越来越多的人的关注,并开始成为知识管理的重要工具。

第二节 机构知识库的定义、特征、功能、类型 及其与其他学术交流模式的区别

一、定义

机构知识库 (Institutional Repository, IR), 又称机构库、机构存储库、机构仓储、机构资料库、机构典藏等。目前,国内外许多学者从不同的角度对机构知识库进行了定义,归纳起来有如下几种。

(一) 从资源的角度进行定义

2002 年, SPARC (The Scholarly Publishing and Academic Resources Coalition, 学术出版和学术资源联盟) 资深顾问 Raym

Crow 最早对机构知识库进行了定义。他认为,机构知识库是获取和保存一个或多个大学智力产出的数字化集合。^①

吴建中^②认为,机构知识库是指收集并保存单个或数个大学共同体知识资源的知识库。

赵继海认为,机构知识库是一个大学或研究机构通过网络来收集、保存、管理、检索和利用其学术资源的数据库。

上述定义仅从资源的角度来对机构知识库进行定义,却忽略了机构知识库中的绝大部分资源是可免费获取的,因为可免费获取是开放存取的一个重要特征。

(二) 从服务的角度进行定义

2003 年, CNI 的常务董事 Clifford A. Lynch^③ 认为, 机构知识库是大学为其员工提供的用于管理和传播机构成员所创作的数字化产品的一系列服务。张智雄和林颖认为, 机构知识库是一系列收集、组织、存储、管理机构内的科学数据和其他资源, 实现机构领域内数字内容和资产的保存与管理、访问和平等使用及共享的一系列机制与服务。

随着实践的发展, 研究者从服务的角度进行定义, 对机构知识库的内涵有了进一步的认识。

(三) 从数字资源和服务的集合角度进行定义

2005 年, 英国 JISC 认为, 机构知识库是建立在个人、部门、机构、全国、地区或联盟基础上的内容存储管理系统, 并向特定的群体提供数字资源内容的服务, 能支持学术、教育和研究活

① Crow R. The Case for Institutional Repositories; SPARC Position Paper. [2010 - 03 - 04]. [http://www.arl.org/sparc/bm/doc/ir final release 102. pdf](http://www.arl.org/sparc/bm/doc/ir%20final%20release%20102.pdf).

② 吴建中:《图书馆 VS 机构库——图书馆战略发展的再思考》, 载《中国图书馆学报》2004 年第 5 期, 第 5~8 页。

③ Lynch C. A. Institutional Repositories; Essential for Scholarship in the Digital Age. *Libraries and Academy*, 2003 (2): 327 - 336.

动。该定义较好地把机构知识库资源与服务的属性进行了结合,但没有充分提示出机构知识库的知识属性。

迄今为止,学术界对机构知识库尚无明确的定义。从机构知识库的本质来看,机构知识库是通过网络对本机构成员的数字知识产品进行捕获、收集、存储、整理、组织、加工和管理,并进行开放存取,以实现对这些数字对象的永久保存和广泛传播为目标的一种知识服务系统。^①因此,我们可以这样认为,机构知识库是学术机构对本机构成员智力产品进行捕获、存储、管理、保存并提供开放利用的知识传播与知识服务系统。它能不同程度地实现数字资源的开放存取和长期保存,促进学术交流,提升机构学术地位的目标等。

二、基本特征

作为一种新型的学术交流模式,机构知识库主要具有以下基本特征。

(一) 主体的特定性

机构知识库的构建主体和收藏的地域界限是特定的机构,其建立和运行均以机构为轴心。机构可以是实体的,也可以是虚拟的,如 e-Science 环境下的虚拟研究实验室等。

(二) 内容的学术性

机构知识库主要收藏本机构的学术性资源,所收集的学术性资源不仅包括常见的研究性学术资源,还包括大量内容复杂、形式多样的且难以通过正式渠道公开发行的灰色文献,如预印本、工作总结、学位论文、研究技术报告、会议文献、技术文献、统计分析、调查报告、教学资源等。由于这些学术资源可反映一个

^① 邓君、毕强:《国内机构知识库研究进展》,载《图书与情报》2007年第5期,第37~42页。