

●赵继海

机构知识库:数字图书馆发展的新领域

摘要 机构知识库是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。它具有学术传播、电子出版、长期保存、知识管理、促进教育、科研评价、共享利用等功能。建设机构知识库,首先要了解其构建过程,定义其功能并制定构建计划,然后组织工作团队、选择技术平台、进入系统启动和运行。目前,机构知识库的发展遇到一些问题和挑战,但其前景广阔。参考文献 8。

关键词 机构知识库 数字图书馆 原生数字化资料 开放利用

分类号 G253

ABSTRACT The author introduces the definition, functions, and development processes of institutional repository, analyzes some problems in this field and proposes future prospects. 8 refs.

KEY WORDS Institutional repository. Digital library. Born-digital materials. Open access.

CLASS NUMBER G253

1 引言

随着计算机、网络和数字化技术的迅速发展,在大学和研究机构,教师与研究人员普遍运用计算机来处理各种信息,从而使原生数字化资料(Born-digital material)急剧增长。这些原生数字化资料包括已发表或未发表的期刊论文、会议论文、专著、学位论文、科技报告、通讯、文件、教学课件、软件、数据库、PPT、声像资料、动画、图片等等,是大学和研究机构的重要学术资源。它们通常分散保存于个人或院系、研究所(室)的计算机上。这种保存方式,既不利于共享利用,也极易因各种原因而丢失。同时,计算机硬、软件发展迅速,分散保存的数字资源,在若干年之后,可能不为更新后的硬、软件所支持。如何长期保存、有效管理和充分利用这些学术资源,给大学和研究机构图书馆提出了新的挑战。

另一方面,传统的印刷型书刊大幅度涨价,使得大学和研究机构图书馆采购印刷型书刊的成本大大上升。在过去的近 20 年间,国内外书刊特别是学术期刊的涨价情况十分严重。据北美研究图书馆协会(ARL)统计,从 1986 年到 2004 年,北美研究型大学图书馆订购的期刊年均涨价达 9.89%,单种期刊的平均价格从 1986 年的 89.77 美元,上涨到 2004 年的 258.73 美元,涨了 1.88 倍。同期图书的年均涨价为 4.05%,单价从平均每册 28.99 美元上涨到 51.19 美元,涨了 76.6%。而图书馆期刊采购经费同期年均增长达 14.37%,平均每个馆从 149.7 万

美元,提高到 558 万美元,采购的期刊种数平均为 22595 种;图书采购经费年均增加 3.32%,从馆均 111.9 万美元,提高到 182.4 万美元,采购的图书册数为 29633 册^[1]。我国无论是进口书刊,还是国内出版书刊,由于各种因素的综合作用,其涨价情况远比发达国家严重。印刷型文献的大幅涨价,使大学和研究机构的文献经费负担日益沉重。学术信息主要是由大学和研究机构产生的,而大学和研究机构却需要向出版商支付越来越高昂的费用来采购这些信息,这显然是不合理的。学术传播方式上的变革,将是大学和研究机构应对上述挑战的必然要求。

既然大学和研究机构是学术信息的产生源,而且这些学术信息又分散存储在研究人员和师生员工、院系或研究所(室)的计算机上,那么,大学和研究机构便可以将这些分散存储的学术信息集中起来,构成所谓的机构知识库(Institutional Repository,简称 IR),再通过校园网甚至通过校际之间的协议,开放利用(Open Access,简称 OA)这些学术信息资源。因此,原生数字化资料急剧膨胀以及出版界对学术书刊长期大幅度涨价,催生了 IR 和 OA 理念的产生,而网络存储价格的持续降低,开放标准如 OAI-PMH 等的确立,则使 IR 技术得以快速发展并被一些发达国家的大学图书馆所采用,成为当前国际数字图书馆发展的新兴和热点领域^[2]。迄今为止国内详尽论述 IR 理念与技术发展的报道不多,鉴于 IR 在数字资源长期保存、开发利用中的重要作用,因而,有必要跟踪发达国家数字图书馆的这一发展领域。

2 IR 的基本理念

关于机构知识库,国内最近有少量介绍,但所用的名称不一,有机构典藏、机构仓储、机构库等不同的译法^[3]。本人认为,将 Institutional Repository 译成机构知识库更适合一些,使得普通用户在字面上就可以理解。这对于一个面向众多用户的服务系统是很重要的。而用仓储、典藏这些名称,前者系计算机专业术语,后者系图书馆学术语,都太过专业化。机构库的译法又不够专指,还可能产生普通仓库的歧义。况且,Repository 一词在现代英语中本身就有知识库、智囊团的意思。同理,本文把 Open Access 译为开放利用,而国内目前对这个概念有开放访问、开放获取、开放存取等不同译法。

Clifford Lynch 最早对大学的机构知识库下了定义:一个大学的机构知识库是学校为师生员工提供的一套服务系统,用于管理和发布由其所产生的数字化资料^[4]。

简单地说,IR 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。我们可以从七个方面的要素来理解它:第一,它是一个基于网络的数据库;第二,它存放的是学术资源;第三,它基于一个大学或研究机构,即由一个特定的机构所界定的知识库,而不是基于某个学科的数据库;第四,它的数据记录具有累积性和长期性;第五,它具有开放和互操作的性质,通过 OAI 兼容软件实现,因而是学术资源开发利用的工具;第六,它用来进行学术资源的收集、存储和发布,是一个大学或研究机构学术传播过程的组成部分;第七,它是数字资源长期保存的工具。

IR 保存、管理与利用的数字化内容包括:研究人员和教师已被学术期刊录用但未发表的论文预印本、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。

上述数字化内容可以文本、图像、音频、视频等不同格式在 IR 系统中保存与利用。

3 IR 的主要功能

IR 作为一个大学或研究机构内部的知识和信息存储与利用的枢纽,其主要功能有:

(1) 学术传播。学术传播是建立 IR 的初衷之一。根据 Lynch 的观点,学术传播(Scholarly commun-

nication)的概念要比学术出版(Scholarly publishing)宽泛得多,后者是前者的组成部分之一^[5]。因此,IR 不能简单地理解为取代学术出版,它将原来不作为正式出版的各种知识、信息通过网络发布,而且具有知识、信息集成的功能。

(2) 电子出版。IR 将各种不同类型的知识和信息内容撷取,按照统一的格式(如 pdf)发布,因而具有电子出版(ePublishing)的功能。通过 IR,可以将原先无法进入正式出版途径的大量知识信息得以发布交流。

(3) 长期保存。IR 为大学或研究机构的知识、信息提供一个长期保存的安全场所,从而克服各部门以及个人分散保存信息的弊端和风险。通过长期的积累,IR 将成为一个大学或研究机构的知识宝库。

(4) 知识管理。通过 IR 的集中式数字对象管理和相互链接,存储其中的信息和知识可方便、快捷、高效地为用户所利用。

(5) 促进教育。大学的师生可以利用 IR 丰富的教学信息资源,提高教学效果。学生可通过 IR 的各类课件进行自主学习、作业和考试,教师可通过 IR 交流、完善教学课件和教学方法,科研人员可通过 IR 进行继续教育和进修提高。

(6) 科研评价。IR 可作为对教师和研究人员科研活动的评价工具,可以使大学或研究机构方便地了解他们的研究成果。

(7) 共享利用。IR 是学术资源开发利用的重要工具。不同的大学和研究机构,通过建立共享利用机制,在权利义务平衡的前提下,相互开放学术资源,对于加强校际之间的学术交流、降低学术信息利用成本具有非常重要的作用。

(8) 提高声望。IR 是对外进行学术交流的重要窗口,可以展示一个大学或研究机构的学术成果,提高论著的被引用率,扩大在学术界的影响力。通过构建 IR,图书馆可进一步发挥在学术传播、信息存储和知识服务中的作用,提升图书馆在大学和研究机构中的学术地位。

4 IR 的构建过程

根据欧美、澳大利亚一些大学图书馆的经验,构建 IR 的主要步骤是^[6]:

(1) 了解 IR 构建的过程。主要是通过对其他大学或研究机构已构建的 IR 进行学习、考察和分析,从而明确构建 IR 需要考虑的主要问题及其主要过程。

(2) 定义 IR 的服务功能。即定义 IR 将要为用户(包括内容提交者和使用者,下同)提供什么样的服务功能。首先应当搞清楚拟建立的 IR 将收集哪些类型的内容,机构内的哪些人可以向 IR 存储内容,主要的服务对象是谁等问题。比如,有些大学的 IR 只接受经过同行评议(Peer review)的资料,另一些则允许接收未经评议的资料;有些只接收教师和研究人员的资料,另一些则可接受博士、硕士生的资料等等。

(3) 制定 IR 的构建计划。在制定 IR 的构建计划前,要进行需求分析。可通过正式或非正式的调查,搞清楚所在大学或研究机构当前学术论著、作品的发表和存储方式,计算机网络及其 IT 设备资源状况,院、系、所和师生员工对数字资源存储、利用及其存在问题的认知程度,参与 IR 建设的主要力量等情况。非正式的调查一般通过与学院教师和管理人员的面谈、邮件联系、跟踪 Web 信息发布与服务情况进行;正式的调查需要设计问卷调查表,通过发放调查表或网上调查,分析所获得的各种数据进行判断。

在明确需求的情况下,制定 IR 构建计划。计划一般由开始阶段、发展阶段和成熟阶段三个时期组成,对不同的阶段有不同的建设目标和服务要求。一般开始阶段应将目标限制在比较小和可控的资源内容、用户对象和服务功能上。在取得成果、取得共识后逐步提升建设目标,增加建设内容,扩大用户范围,扩充服务功能。

根据 IR 构建计划,建立一个成本模型(Cost model),确定构建 IR 的工作进程表和时间安排。建立内容提交、发布和维护管理的制度,确定对用户是完全免费还是需要部分收费的政策。制定 IR 的短期目标和长期目标,确定优先的服务功能和服务对象。确定图书馆与大学或研究机构内各单位之间在 IR 构建中的责任。

(4) 组建工作团队。根据 IR 构建计划任务,确定所需的工作人员及其职责,物色相应人员,组建 IR 建设工作团队。这是保证 IR 构建计划完成与服务效果的关键。

在计划阶段,需要物色需求调查与分析、服务模型构建、成本分析、资源评价、计划制定的人员,组成一个 IR 实施计划小组。一般地说,需要图书馆领导参与其中。

在实施阶段,主要是组建用户服务与系统技术两类专业人员,最重要的是物色这两支专业队伍的主管。

用户服务人员主要负责与用户的沟通和技术支持,特别是对向 IR 提供内容的用户的技术支持,其主要职责是:

- ①负责 IR 系统的用户管理,开发 IR 网站。
- ②负责用户培训,包括图书馆员、IR 内容提供者和终端用户的培训,提供电话、在线或邮件的用户支持服务,编写与更新 FAQ。
- ③协调管理 IR 内容的提交,与学科馆员沟通,确定内容新类目的定义和设置。
- ④配合系统管理员,协调历史数据批处理方式输入 IR 系统。
- ⑤负责元数据创建、定制、评价和移植。
- ⑥负责 IR 内容管理和长期保存。
- ⑦负责与院、系、所的沟通,根据用户和图书馆员的反馈意见,向系统开发人员提出 IR 的新功能需求。
- ⑧负责宣传推广 IR 的服务。
- ⑨对学校制定 IR 服务的政策、规范、制度及其相关法律问题提供咨询意见。

⑩用户服务的主管应当参与 IR 的有关决策、计划过程。

要完成这些工作职责,用户服务人员特别是主管需要具备较高的图书馆学专业水平,能熟练运用相关软件工具,尤其需要熟悉 Web 发布工具,具有良好的语言文字功底和沟通能力,具有较广的知识面。

系统技术人员主要负责 IR 系统的技术方面的工作,其主要职责是:

- ①负责 IR 系统各种技术问题的管理。
- ②负责系统技术培训,协调 IR 系统与图书馆已有系统之间的关系。
- ③负责系统维护、Bug 修复,并向系统开发商提供系统改进的意见。
- ④负责系统管理、监视、测试和调试。
- ⑤负责应用程序和中间件的监测、升级。
- ⑥开发系统延伸功能。
- ⑦管理 IR 服务器及其他硬件、合同、文件、记录硬件运行中的问题。
- ⑧参与 IR 项目决策、计划与服务活动。

系统管理员要求具备扎实的计算机专业功底,熟悉图书馆业务,熟悉 Unix 系统,具有 Java 编程的经验,了解 Web 开发技术,熟悉基本的系统管理技术,具有从事开放源代码开发项目的经验。

(5) 选择适用的 IR 技术平台。目前,已经有一批 IR 软件平台可供大学和研究机构选择^[7]。IR 软

件平台可分为开放源代码的免费软件和商用软件两类。对于开放源代码的免费软件,用户可下载这类 IR 软件平台,可检查、定制和改进源代码。这类软件平台有 CDSware、DSpace、EPrints、Fedora、Greenstone 等等,其中应用较广的是 DSpace,它由 MIT 和 HP 合作开发^[8]。商用软件的供应商拥有软件平台的所有权,用户有三种服务模式可供选择:第一,本地运行和管理;第二,本地运行,软件商远程管理;第三,软件商远程运行和管理。不同的服务模式有不同的费用标准。软件商负责源代码的控制和软件的升级,还可代为用户管理数据。这类软件目前有 Open Repository、bepress 等。

(6)宣传推广。教师、科研人员的广泛参与和决策层的人力支持是 IR 建设与服务的前提。因而,IR 工作团队要在一个大学或研究机构内,利用各种媒体,深入院、系、所,广泛宣传建立 IR 对于教学、科研和学术传播的重要意义与作用,普及 IR 知识,以取得各方面的理解、共识、兴趣和支持。

(7)系统启动。当完成了前面几个步骤后,IR 系统就可进入启动阶段。为所选择的软件平台配置相应的服务器及其他硬件设备,对软件平台进行安装和参数设置,定制界面,培训技术人员,创建内容核准工作流(包括接受、编辑、拒绝等等),安装数据内容,就可进行系统的测试和启动。

欧美一些大学在 IR 系统启动阶段,一般先进行试验,挑选积极性较高、学科交叉、数据内容的类型或格式多元的院、系、所作为试验对象。请有关教师和科研人员提交数据内容,对软件平台进行测试,解决内容提交与服务过程中的一些技术问题。同时,对 IR 有关的政策和制度进行试点,为向整个学校全面推广提供经验。

(8)系统运行。经过系统启动阶段的试验,取得经验后,IR 系统就进入了运行和服务阶段。与自动化集成系统和数字图书馆系统一样,IR 系统需要实行 24/7 的不间断运行方式,要考虑数据的备份与恢复,系统的维护,数据的增长,服务范围的延伸和服务对象的扩大,要有定制功能,并逐步朝着校际合作、国际合作的方向发展。

系统运行过程中,内容数据的组织和增长,是 IR 服务最重要的资源基础。如何使一个大学或研究机构的教师、科研人员主动、及时地向 IR 提交内容,是 IR 建设和服务的最大挑战,也是 IR 团队最重要的工作。内容的组织应当与一个大学或研究机构的学术

机构设置以及独特的文化相适应。欧美一些大学通常采用按院、系或研究所、研究中心来组织内容,这是比较方便的方式。当然,也可以像美国堪萨斯大学那样采用混合交叉的三种方式组织内容,即:①通过现存的院系和研究中心组织内容,特别是对研究对象比较确定的院系和研究中心,通常采用这种方式,如神经系统科学系;②通过学科群组织内容,所有的教师和研究人员都可以直接或通过代理提交相关内容,经过图书馆员评审后再在网上发布,对于跨度较大的人文社会科学研究资料的组织,宜采用这种方式,如公共政策研究;③通过建立专门的兴趣小组组织内容,人员跨院系,按研究兴趣组成,如社会学研究。

为了吸引教师和研究人员提交内容,根据欧美大学的一些做法,IR 工作团队需要从几个方面努力:一是持续不断地宣传、推广 IR 的知识与意义。二是 IR 的用户界面一定要有吸引力,要便于使用和提交内容,用户很容易得到在线帮助。三是 IR 工作团队要与学校的各类网站建立良好的合作关系,了解教师、科研人员论著、成果发表状况,并努力充实到 IR 系统。四是制定科学、合理、公平的内容提交和使用政策,保护知识产权。

5 结束语

IR 对于保存数字资源、促进学术传播、开放利用信息资源具有重要的作用,受到了欧美等发达国家和地区大学、研究机构以及 IT 厂商的关注,成为数字图书馆的重要发展领域。但是,欧美一些研究型大学在建立 IR 的实践中遇到了不少的问题和挑战。首先,教师和研究人员参与 IR 内容提交的积极性存在较大差距,获得各级管理部门的支持情况各异。其次,厘清和管理 IR 内容数据的知识产权难度较大。第三,IR 系统收费运行难以得到响应,免费运行需要持续不断的经费投入,成本管理和运行政策陷入两难境地。第四,数字资源长期保存挑战严峻。我国大学和研究机构关于 IR 系统建设还未见实质性起步的报道,但从目前国外发展的势头看,我国构建 IR 系统只是时间问题。上述问题与挑战,将同样存在于我国,而且有可能会以更严重的程度表现出来,在制定 IR 系统建设规划时须作周密考虑。

应当指出,目前即使在发达国家,IR 系统还只是学术传播的非主流模式。互联网的发展,数字化技术的集聚,是不可阻挡的进程。IR 作为数字资源长期保存的工具和开放利用的枢纽地位(转第 50 页)

- Technology, Communications of the ACM, 36(12), 1993
- 2 甘利人.企业信息化建设与管理.北京:北京大学出版社,2001
- 3 Bresnahan, T. F., and M Trajtenberg. General Purpose Technologies; Engines of Growth? Journal of Econometrics, 1995, Vol. 65 (5)
- 5 李翔.企业信息化评估体系初探. <http://www. icad. com. cn>(2005年3月15查阅)
- 6 Hitt L M, Brynjolfsson E. Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value, MIS Quarterly, 20(2), 1996
- 7 约翰·索普. DRM战略领导中心.信息悖论.大连:东北财经大学出版社,2003
- 8 李航.别把“上线”当终点! AMT前沿论丛,2005(22)
- *****
(接第36页)是可以预期的,它作为学术传播的重要模式也终将会为人们所接受。虽然它在未来一定时期内,不会取代有数百年发展历史的学术出版模式,但两种模式共存的局面则是可以预期的。IR与OA相结合,必定会对长期以来涨价幅度过大的学术书刊出版业带来冲击。图书馆担负着知识传承、信息服务、学术传播的责任,无疑应当关注、跟踪IR技术和应用的发展,对其可行性进行评价,并依据内外部环境的变化,及时构建IR服务系统。

致谢:作者受“包兆龙、包玉刚中国留学生奖学金”高级访问学者计划资助,于2005年7~9月赴美进行3个月的访问研究,本文为访问研究的成果之一。特向包氏基金会、美国卡内基·梅隆大学图书馆、匹兹堡大学图书馆等表示衷心感谢!

参考文献

- 1 Association of Research Libraries. ARL Statistics 2003 - 04. <http://www. arl. org/stat/machine/2003 - 04>
- 2 Westrienen G. V. , Lynch C. A. Academic Institutional Re-

- 9 Sanjeev Dewan, Kenneth L. Kraemer. International Dimensions of the Productivity Paradox, 1998, Communications of ACM, Vol. 41(8)
- 10 Brynjolfsson E, Hitt L. M. , and Yang S. (2002a). Intangible Assets: How the Interaction of Computer and Organization Structure Affect Stock Market Valuations. Brookings Papers on Economic Activity: Macroeconomics(1), 2002

陈远 武汉大学信息管理学院副教授。通信地址:武汉市。邮编430072。
彭珍 武汉大学信息管理学院硕士研究生。通信地址同上。

(来稿时间:2005-05-08)

- pository: deployment status in 13 nations as of mid-2005. D-Lib Magazine, 2005, 11 (9). <http://dlib. org/dlib/september05/westrienen/09westrienen. html>
- 3 董文鸳,袁顺.聚集学术机构知识的中心:机构库(Istitutional Repository)探析.图书馆杂志,2005(8)
- 4,5 Lynch C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. ARL Bimonthly Report, 2003, (226): 1 - 7
- 6 Barton M. R, Waters M. M. Creating an institutional repository: LEADIRS workbook. <http://www. dspace. org/implement/leadirs. pdf>
- 7 Budapest Open Access Initiative. Guide to institutional repository software. <http://www. soros. org/openaccess/software/>
- 8 MIT's DSpace experience: a case study. <http://dspace. org/implement/case-study. pdf>

赵继海 浙江大学图书馆副馆长,研究馆员。通信地址:杭州市浙大路38号。邮编310027。
(来稿时间:2005-10-12)