

图书馆科研数据管理与服务模式探讨^{*}

李晓辉

摘 要 基于国外图书馆开展科研数据管理与服务的实践,从技术支撑、科研数据组织方式、科研数据服务方式、基于科研数据的用户信息素养教育、相关机构与人员配置等方面探讨了图书馆进行科研数据管理与服务的模式。技术支撑包括基础设施构建和软件工具建设;科研数据组织方式包括机构仓储方式和数据门户方式;科研数据服务包括数据获取服务和数据分析服务;科研数据管理应纳入到用户信息素养教育中,对科研人员进行科研数据管理意识、知识和技能的培训;科研数据管理与服务需要图书馆相关机构和人员配置予以保障。图2。表1。参考文献25。

关键词 图书馆 科研数据 管理与服务 信息素养教育

分类号 G250

ABSTRACT Based on the research data management experience in foreign libraries, this article explores the pattern of research data management and service in libraries from five aspects: the technology support, data organization, data services, user information literacy education and department and human resource configuration. Technology support includes infrastructure construction and soft tool design; the ways of research data organization include institutional repositories and data portal; research data service includes data access and data analysis. Research data management should be brought into users' information literacy education in order to provide researchers with necessary training on awareness, knowledge and skills regarding research data management. Research data management and service need to be guaranteed by relevant departments and personnel. 2 figs. 1 tab. 25 refs.

KEY WORDS Library. Research data. Management and service. Information literacy education.

CLASS NUMBER G250

随着海量数据时代的来临,学术交流环境和科研环境的变化,科学研究正在向数据密集(data intensive)型科研转变^[1]。这种转变使图书馆面临新的任务(本文中图书馆指大学图书馆和研究图书馆),即图书馆如何获取新技能、如何调整角色为 e-Science 和 e-Research 提供数据支撑?如何对颇具科研价值的数据进行管理和服务?澳大利亚、美国、英国的一些大学已把科研数据管理作为图书馆服务的一项内容。国内大学图书馆的科研数据管理尚未系统化,只是在培训过程中涉及一些相关内容,还没有提高到科研数据管理的层面,也没有完整的规划。本文通过总结国外图书馆科研数据管理的实践,挖掘图书馆进行科研数据管理与服务的共

性,探讨图书馆科研数据管理与服务的模式。

1 科研数据管理与服务的内涵

科研数据(Research Data)是指数字形式的研究数据,包括在研究过程中产生的能存贮在计算机上的任何数据,也包括能转换成数字形式的非数字形式数据。如传感器读取的数据、遥感勘测数据、调研结果、神经图像、实验数据及来自测试模型的仿真数据等^[2]。格式有文本的、数值型的、多媒体的,也包括模型和软件等。有效的科研数据管理具有保护数据免于丢失、提高数据曝光度,传播和出版成果、实现数据共享、对科学质疑公开、鼓励观点的多样性、节约

^{*} 本文系2010年度教育部人文社会科学研究青年基金项目“图书馆信息资源平台社会共享研究”(编号:10YJC870025)研究成果之一。

科研成本、完成研究资助方的要求等诸多意义。

图书馆开展科研数据管理与服务,既是用户的外在需求,也是图书馆自身发展的内在需要,是一种跨界服务、嵌入服务、动态服务。跨界服务是指科研数据服务超越了图书馆传统的服务边界,跨越了文献信息服务,深入到科研数据;嵌入服务是说科研数据管理与服务需融入到用户一线,嵌入到用户科研环境和科研过程,提供个性化服务,方便用户集成化管理各种科研数据;动态服务是因为科研数据服务是系统化工程,需要与用户建立长期的沟通与协作,在与用户的反馈中不断调整策略,是一个连续的

动态过程。这是图书馆科研数据服务的特点。

2 图书馆科研数据管理与服务的模式

图书馆科研数据管理与服务是一个系统工程,需要图书馆及大学内多个部门协调配合,才能完成科研数据管理生命周期的所有任务。根据国外图书馆的实践,本文从技术支撑、资源建设、服务方式、培训内容和方式、相关部门与人员配置五个方面,构建图书馆科研数据管理与服务模式(见图1)。

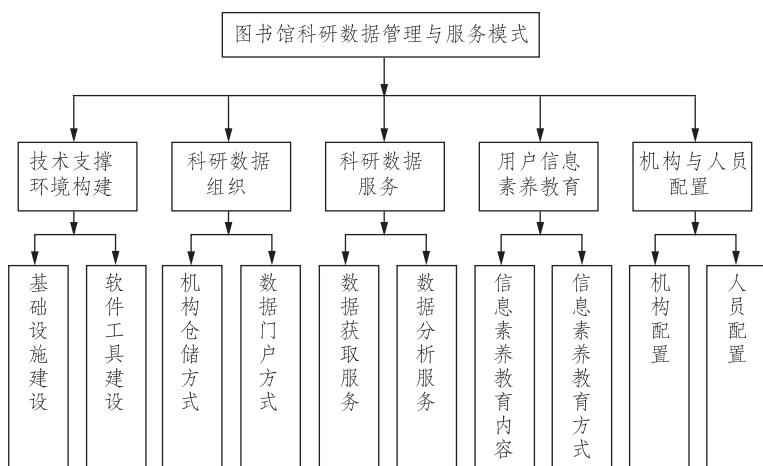


图1 图书馆科研数据管理与服务模式框架

2.1 科研数据管理与服务的技术支撑环境构建

对大量的、复杂的数字化形式的科研数据进行管理与服务,离不开信息技术的支撑,需要依托一定的基础设施和软件工具。主要是大型存储设备,高性能服务器及服务终端搭建,软件系统等。图书馆是大学基础设施的重要组成部分,但图书馆不能独立完成科研数据管理与服务生命周期的所有任务,需要与校内外的各种机构合作,充分共享现有的设施和技术,共同构建科研数据管理与服务所需的技术环境。

2.1.1 基础设施构建

(1)图书馆自建或与校园内其他部门共同

建设相关基础设施。图书馆独立建设数据仓储并对其进行维护与管理,目前这种形式并不多,这并不仅仅是图书馆的技术能力问题,在开放、共享的理念下,图书馆更多地趋向于合作方式,重点利用学校建立的数据或机构仓储,利用已有的基础设施,与校园内其他部门合作建设科研数据管理的硬件环境。如美国明尼苏达大学图书馆与明尼苏达信息技术办公室、明尼苏达机构库、明尼苏达大学塞博基础设施联盟合作^[3]。澳洲国立大学建立的超级计算机设备(Australian National University Supercomputing Facility,ANUSF),提供了大量项目范围内的数据存储服务^[4]。在校园内其他机构已经建立数

据仓储的条件下,图书馆不再需要存储数据,不需要搭建硬件环境,需要的仅仅是合作的理念。

(2)图书馆与校园外的各种联盟及项目机构合作,充分利用外部仓储设备。政府和一些组织资助的项目也建立了不同学科的数据仓储,为国内机构提供共同服务,使科研数据存储超越了图书馆,超越了某一个单一机构。图书馆这时的任务是向研究者提供相关信息与帮助,使其了解这些仓储,并帮助用户利用这些重要资源。在技术上侧重科研环境建设,构建数据门户,做好数据导航。如澳大利亚社会科学数据存档(Australian Social Science Data Archive, ASSDA)是全国领先的澳洲大学联盟,目的是收集和保存计算机可读的与社会、政治和经济事件相关的数据,并使这些数据将来可获取^[5]。这些机构都是图书馆的合作对象。

(3)利用云计算共享基础设施。云计算会促进 e-Science 发展,通过云远程使用资源,图书馆不需要自己购买硬件设施,不需要建设基础设施环境,大型设备尽可在网上利用,图书馆可以将数据存储的云存储服务提供的服务器中,按需申请,按时付费。如亚马逊的 S3, 微软的 Azure 等。目前云计算在图书馆的应用并不多,OCLC 启动的“将图书馆管理服务推向 Web 级的战略”是图书馆界接受云服务的重要标志性事件^[6]。以后会有越来越多的图书馆试水云计算,通过云计算降低管理与服务的成本。

2.1.2 软件工具支撑

科研数据管理过程中,不同阶段需要不同的软件支持,常用软件和具体功能见表 1(以澳洲国立大学图书馆为例)^[2]。软件服务的原则是充分利用开源软件,购买被广泛应用的商业软件,减少自主开发软件。

目前图书馆基本采用前两种形式。利用开源软件,或代表机构购买软件使用权。如澳洲国立大学代表机构获得 Endnote 软件的使用协议,允许学校内的研究人员下载使用。同时,图书馆把各种开源软件与购买的软件集成在一起,提供软件和工具服务。如墨尔本大学图书馆提供多种类型软件的免费下载,并提供软件安装、使用的相关说明和帮助。软件工具分为

五类:①图书馆博客:用以跟踪图书馆试用数据库,软件更新及新软件情况;②参考书目管理工具:帮助研究人员记录、组织、使用参考文献;③检索工具;④链接到在线的文章或资源工具:链接到在线的文章或资源,保留喜好的资源目录;⑤教职员工具:提供已订阅书目的在线电子版阅读等^[7]。英国爱丁堡大学图书馆也提供一系列相关免费软件的下载服务,包括版本控制和加密软件^[8]。

表 1 科研数据管理的相关软件

序号	软件名称	软件功能
1	Endnote、JabRef	参考书目管理
2	Pebble、Alliance	文件传输和远程获取
3	Winscp	同步
4	Alliance	合作
5	TortoiseSVN	版本控制

无论用哪种形式构建基础设施,图书馆都要致力于建设便于用户科研数据管理与利用的环境,把校园内外的技术设施与用户需求整合到一起,构建用户科研数据管理平台,并把其嵌入到用户的科研环境、科研过程,满足科研人员信息交流、数据共享的需求,同时满足数据存储、管理、长期保存、复用的需要。

2.2 科研数据组织

经过组织的资源才能被用户充分利用和共享。图书馆对科研数据的组织按其存储地点分为两种情况:一种是存储在图书馆内的数据,以机构仓储方式组织;另一种是存储在图书馆外的数据,以数据门户方式进行组织。

2.2.1 机构仓储方式

机构仓储最初的设想是保存机构成员的研究成果,并提供出版机会,既有存储的功能,又有检索和服务的功能。研究人员的科研数据除保存在相关学科库以外,机构仓储是另一重要选择。通过对存储在机构库中的数据所做的一系列工作,减轻了数字过时的风险,增加了其长期研究的价值,数据存储可在可信赖的仓储可以

让更大范围的人共享。目前许多大学和研究机构都建立了自己的机构仓储,有的还建有专门的数据仓储,如美国哈佛—麻省理工数据中心(Harvard-MIT Data Center, HMDC)是两校联合建立的数据仓储,供教师存储并检索数据^[9];霍普金斯大学(John Hopkins)的谢里登学院(Sheridan)图书馆,建有数字研究和长期保存中心(Digital Research and Curation Center),中心强调发展自动化工具、系统软件以减少成本,尤其重视对大规模的原生科研数据的收集和长期保存^[10]。

2.2.2 数据门户方式

对于存储在机构外的科研数据,图书馆建立科研数据门户,提供导航服务。这种形式下数据没有存储在图书馆,但图书馆可以通过链接方式链接到其他的存储中心,建立科研数据知识组织环境,把与数据管理相关的内容集成到一个系统内,统一进行管理。如爱丁堡大学图书馆建立了数据仓库(Data Library),目的是支持用户发现、获取、使用和管理科研数据,使数据能够用于分析、教学和课程作业。门户下分五个专题,分别是:①发现数据:浏览拥有的数据,并发现网上免费数据;②获取数据:获得国家各种科研数据和图书馆收藏数据;③使用数据:通过用户指南和在线指导,帮助用户使用科研数据;④管理数据:指导用户管理、共享和保存数据;⑤数据仓储服务:对如何使用爱丁堡大学的数据共享仓储(Edinburgh DataShare repository)服务并共享科研数据进行指导^[11]。美国哥伦比亚大学图书馆也建有数据门户(data-gate),通过门户主要提供适用于GIS(地理信息系统)绘图软件的空间数据,适用于各种统计软件的数据^[12]。

2.3 科研数据服务

数据管理是为了将来可以复用。图书馆科研数据服务按服务中需要的智力程度分为直接数据获取服务和数据分析服务。前者是基础,后者是未来趋势。

2.3.1 数据获取服务

直接为用户提供数据获取服务是图书馆的

一种基本数据服务方式。首先是为用户提供存储在图书馆的数据服务。如大英图书馆接受了威尔士方言调查(Survey of Anglo-Welsh Dialects, SAWD)的报告,调查数据的音频文件副本2008年已捐赠给大英图书馆,并通过图书馆的声音服务器系统,以MP3的格式为用户提供获取服务。通过这种方式,使珍贵的研究原始数据获得更大的曝光率和利用率^[13]。其次,存储在图书馆外的资源,图书馆可以通过其构建的门户为用户提供导航服务,方便用户获取数据。如哥伦比亚大学通过其门户为用户提供如下服务:识别数字和空间数据;从数据源中捕获或抽取数据;把数据转换成适合研究和分析的格式^[12]。爱丁堡大学通过其门户也为用户提供发现数据和获取数据服务^[11]。

直接获取数据的服务与图书馆的其他资源服务(如电子数据库)相类似,是图书馆开展科研数据服务的第一步。在此基础上,图书馆可进一步展开对科研数据深层分析服务。

2.3.2 数据分析服务

未来图书馆科研数据服务趋势是服务中附加更多智力活动,进行数据分析,把科研数据与相关科研文献相关联,帮助用户发现相关数据,获取数据等。根据现有资料,尚未有图书馆提供相关服务。但一些机构仓储在存储数据的同时,提供了类似的服务。如哈佛大学的“数据组诗网络”(Dataverse Network),项目包括科研数据出版、共享、参考、抽取和分析各个方面,为大学或其他机构提供数据出版系统的全部解决方案,并提供数据分析服务,产生正式的数据引用(data citation),通过SPSS和STATA分析数据表,并可下载结果,使数据服务层次不断深化^[14]。目前可提供数据分析的机构尚少,这是图书馆科研数据服务的方向。

2.4 基于科研数据管理的用户信息素养教育

信息素养教育是图书馆的一项重要工作。国外一些大学已将科研数据管理纳入到信息素养教育的内容中,对科研人员进行科研数据管理意识、知识和技能的培训。早在2008年,澳大利亚收藏在线研究(Online Research Collections

Australia, APSR)就资助澳洲国立大学的信息素养规划,形成一个关于研究生数据管理的培训规划^[15]。

2.4.1 信息素养教育内容

根据国外图书馆的实践,可将科研数据管理的内容分为三部分:组织、管理、存档与共享^[2](见图2)。对于用户的信息素养教育主要围绕以下内容进行。

(1)科研数据组织。包括:参考书目管理;文件传输和远程获取,即使文件可以远程传输和修改;同步,即在不同的地点使研究同步;多人合作书写与修改;版本控制。

(2)科研数据管理。包括:数据备份;数据

验证和鉴定,即通过制订策略和实践方法,保证数据的准确和真实;文档记录,即把技术报告、实验记录、数据收集创作的方法、数据分析方法记录下来,保证后续工作;获取控制,即对数据的获取级别进行限制;数据安全,即使用各种软件保证文件的安全。

(3)科研数据存档和共享。包括:共享方法,即把数据存档在某一仓储;版权和许可,即研究者对自己的科研数据授权,规定他人的使用权限;制订使用格式和标准,即确定文档使用的标准和格式等;确定获取限制;元数据管理;数据存档。

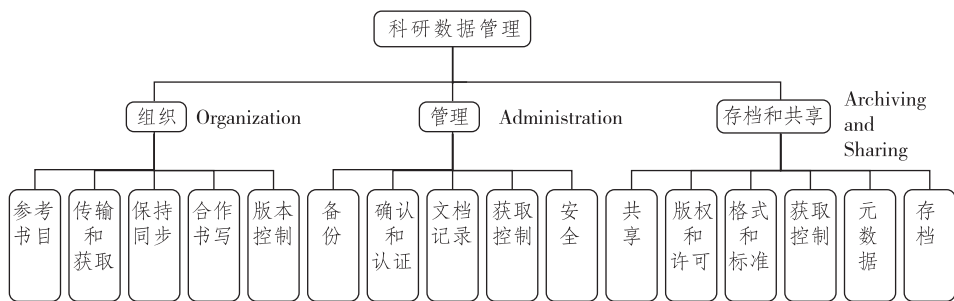


图2 科研数据管理培训内容

2.4.2 信息素养教育方式

目前图书馆开展的基于科研数据管理的用户信息素养教育方式分为两种:一种是普适性的信息素养教育;另一种是针对某一主题领域的科研数据进行的学科化信息素养教育。

(1)普适性的信息素养教育。这种方式面对所有用户,目的是让用户了解科研数据管理的目的、意义和要求;熟悉数据管理的一般方法和术语;培养数据管理实践所需技能;关注与科研数据管理的相关政策法律;制定合理的数据管理计划。又可分专题讨论(workshop)、在线学习(online)二种。目前国外许多大学都组织了相关形式的科研数据培训工作,如澳洲国立大学为研究者和本科生提供数据管理计划的课程。美国麻省理工大学图书馆把科研数据管理纳入其信息素养教育的内容^[16],图书馆还制定统一的数据管理规划清单和模板,同时提供详

细的数据管理手册,数据管理课件供下载利用^[17],用户可以选择参加专题讨论或在线自主学习,此外,培训的形式和内容也在不断完善,成为学生课程内容的一部分。爱丁堡大学图书馆2010年9月启动了历时一年的科研数据管理项目,目标是开发在线学习教材。这些教材将反映科研数据管理的最佳实践、就科研数据管理对重要学者的视频采访,及软件使用方面的经验等,最终结果与研究生课程相结合,并可在在线获得^[18]。美国斯坦福大学图书馆将科研数据培训与虚拟校园(Virtual Campus)的课程整合在一起,供学生注册后免费学习^[19]。

(2)学科化的信息素养教育。针对某一领域主题,图书馆建立科研数据管理培训实验室,不仅要配备相关基础设施、软件、人员,还要提供一定的物理空间,为用户的科研数据培训提供一站式个性化服务。如在麻省理工学院,提

供生物信息学培训,学习如何用解释分析系统(BIOBASE'S)、解释基因表达、切碎数据^[20];地理信息系统服务项目(Geography Information System GIS)实验室由图书馆、信息服务和技术部门、教育创新技术办公室合作建设,支持数据收集,训练和协助使用GIS,但不做项目^[21]。斯坦福大学图书馆的社会科学数据和软件(Social Science Data and Software, SSDS)为师生提供社会科学数据存储与检索以及相关定量统计和定性分析软件的选择和使用培训,并配备具体的实验室^[22]。

2.5 相关组织机构和人员配置

2.5.1 图书馆机构配置

科研数据管理对图书馆来说还是新事物,应对复杂的科研数据管理工作,需要组织机构和人员配置方面的相关配合。

首先,成立专门的协助部门。科研人员在实践中会遇到各种问题,图书馆需要为此提供相关的帮助。如麻省理工学院图书馆成立免费的元数据服务小组,为学校的研究项目提供帮助,为元数据应用提供专家意见,通过服务为科研项目节省时间^[23]。其次,成立专门的管理部门,墨尔本大学图书馆在2010年规划中提出建立科研数据管理办公室,以发挥图书馆在科研数据管理中的作用^[24]。相关的组织机构配置是图书馆顺利进行科研数据管理的保障。

2.5.2 图书馆人员配置

图书馆增添科研数据管理的新内容,对图书馆员能力和知识结构提出了更高要求,图书馆要为馆员创造条件使其有能力参与数据管理。墨尔本大学图书馆在图书馆学员纲要中提到雇佣数据专家馆员(Data Specialist Librarian)^[24]。此外,国外对参与科研数据管理的馆员的称谓还有:数据人文主义者(Data Humanist)、数据科学家、数据研究科学家(Data Research Scientist)、数据服务馆员(Data Services Librarian)等^[25]。从这些称谓的变化可以看出图书馆员角色的扩展。图书馆可对馆员进行培训,也可直接引入所需人才,使馆员具备科研数据管理与服务所需的专业知识与技能。

3 结语

把科研数据作为图书馆管理与服务新内容的实践刚刚起步,图书馆应积极调整自己的角色,努力发挥图书馆在科研数据管理中的作用。技术支撑环境、对数据资源进行科学合理的组织、适当的服务方式、基于科研数据管理的信息素养教育、相关组织及人员配置构成科研数据管理与服务的五大要素,贯穿于科研数据管理的全过程。几大要素紧密结合,才能使图书馆的科研数据管理和服务工作有序展开,使图书馆的工作紧跟信息环境的变化,紧跟用户的需求,为新的研究范式下的科研活动提供数据支撑。未来几年,图书馆与校园内外各种机构和科研人员的合作将更加广泛,科研数据管理与服务在图书馆将变得更加普遍。

参考文献:

- [1] Hey T, TanSley S, Tolle K. The fourth paradigm-Data-intensive scientific discovery[OL]. [2010-09-14]. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf.
- [2] ANU data management manual: Managing digital research data at the Australian National University [OL]. [2010-09-01]. http://ilp.anu.edu.au/dm/ANU_DM_Manual_v1.03.pdf.
- [3] Delserone L M. At the watershed: Preparing for research data management and stewardship at the University of Minnesota Libraries [J]. Library Trends, 2008, 57(2): 202-210.
- [4] ANU supercomputer facility[OL]. [2010-09-01]. <http://anusf.anu.edu.au/>.
- [5] The Australian Social Science Data Archive (ASSDA) [OL]. [2010-09-01]. <http://assda.anu.edu.au/about.html>.
- [6] OCLC 宣布将图书馆管理服务移到“网络规模”的策略[OL]. [2010-09-14]. <http://www.oclc.org/asiapacific/zhtw/news/releases/200927.htm>.
- [7] Software and tools of the university of Melbourne

- library[OL]. [2010-09-01]. http://www.library.unimelb.edu.au/help_yourself/tools.
- [8] The university of Edinburgh information services; Training, advice & support for data management [OL]. [2010-09-01]. <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/training-advice-support/tools>.
- [9] Harvard-MIT Data Center[OL]. [2010-10-31]. <http://hmdc.harvard.edu/>.
- [10] Digital Research and Curation Center[OL]. [2010-10-31]. <http://ldp.library.jhu.edu/dke>.
- [11] Data library [OL]. [2010-09-06]. <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/overview>.
- [12] DataGate_Columbia's gateways to numeric data[OL]. [2010-10-31]. <http://www.columbia.edu/cu/lweb/indiv/dssc/eds/dgate.html>.
- [13] British Library Sound Archive welcomes the Survey of Anglo-Welsh Dialects[OL]. [2010-09-01]. <http://www.bl.uk/news/2008/pressrelease20080723a.html>.
- [14] An Open-source application for publishing, citing and discovering research data[OL]. [2010-09-01]. <http://thedata.org/book/features>.
- [15] Data management planning[OL]. [2010-10-31]. http://www.apsr.edu.au/orca/data_management_planning.htm.
- [16] MIT libraries data management and publishing[OL]. [2010-09-01]. <http://libraries.mit.edu/guides/subjects/data-management/index.html>.
- [17] ANU information literacy program[OL]. [2010-09-01]. <http://ilp.anu.edu.au/dm>.
- [18] Research Data MANTRA project[OL]. [2010-10-31]. <http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/mantra/about>.
- [19] ESRI's virtual campus[OL]. [2010-10-31]. <http://training.esri.com/gateway/index.cfm>.
- [20] BIOBASE[OL]. [2010-09-01]. <http://libguides.mit.edu/content.php?pid=14149&sid=171340>.
- [21] Geographic information systems services@mit[OL]. [2010-09-01]. <http://libraries.mit.edu/gis/>.
- [22] Social Science Data and Software (SSDS)[OL]. [2010-10-31]. <http://www.stanford.edu/group/ssds/cgi-bin/drupal/content/about-ssds>.
- [23] Metadata Services MIT Libraries[OL]. [2010-09-01]. <http://libraries.mit.edu/metadata/>.
- [24] Melbourne University Library Annual Plan 2010[OL]. [2010-09-01]. <http://www.library.unimelb.edu.au/?a=289993>.
- [25] Shreeves S L, Cragin M H. Introduction; Institutional repositories; Current state and future [J/OL]. Library trends, 2008, fall;89-97 [2010-09-01]. <http://www.ideals.illinois.edu/bitstream/handle/2142/10679/shreevescragin.pdf?sequence=2>.

李晓辉 哈尔滨师范大学图书馆副研究馆员，在读博士生。通讯地址：哈尔滨市南岗区和兴路50号。邮编：150080。

(收稿日期:2010-09-19;修回日期:2010-10-31)