

●彭斐章 邹瑾

网络环境下的信息存取与目录学创新^{*}

摘 要 从目前书目数据的开放存取、开放存取资源目录、书目数据通用开放标准、网络书目情报服务、目录强化与信息存取,以及开放源代码、书目软件等技术和工作现状,可以看到目录学理论的新发展和书目工作的不断创新。对这些方面的分析研究有利于目录学的发展。参考文献 31。

关键词 网络信息 开放存取 书目数据开放 目录学研究 书目服务

分类号 G257

ABSTRACT The authors analyze the open access of bibliographical records, directories of open access resources, common open standards of bibliographical records, online bibliographical information services, bibliographical enrichment and information access, open source code software for bibliographical records, etc., and discuss the recent developments of bibliographical studies and innovations in bibliographical work. 31 refs.

KEY WORDS Network information. Open access. Open bibliographical records.

Bibliographical studies. Bibliographical services.

CLASS NUMBER G257

近年来,随着开放存取(open access)、公共获取(public access)等思想及活动的涌现,网络信息存取的开放、公平和自由的特点得到更加充分的展示。正是由于网络信息存取的增加,关于信息资源的知识和效用信息在信息存取中变得极为重要,成为网络环境下书目情报工作和目录学研究的重点。

目录学是致用之学,同时,目录学也一贯具有鲜明的时代特征。目录学一方面要为开放的信息环境服务,一方面也要积极调整开放策略。不仅要推进用户书目情报的存取,也要通过自身的强化拓展数字信息的存取。网络环境下,目录及书目情报工作实践的发展不断为目录学理论的充实和创新扩展空间。

1 书目数据的开放存取

“开放存取”是基于“自由、开放与共享”的理念,依托网络技术的一种学术出版和科学信息交流的全新模式^[1]。2001年,“布达佩斯开放存取先导计划”(BOAI)给出的“开放存取”的定义有如下表述:“对某文献的‘开放存取’即意味着它在 Internet 公共领域里可以被免费获取,并允许任何用户阅读、下载、复制、传递、打印、搜索、超链接该文献,也允许用户为之建立索引,用作软件的输入数据或其他任何合法用途。用户在使用该文献时不受财力、法律或技术的限

制,而只需在获取时保持文献的完整性,对其复制和传递的唯一限制,就是要保证作者有权控制其作品的完整性,同时在使用作者作品时注明相应的引用信息……”^[2]。

开放存取学术资源数据库的范围包括商业索引数据库的开放存取版、开放存取索引-文摘数据库和开放存取全文数据库^[3]。虽然全文数据库所占比重越来越大,但索引、文摘等书目数据库仍不失为一种重要的参考数据类型。书目资源的开放与共享是书目情报需求全球化发展的需要,学术资源开放存取为网络环境下书目情报开发利用和书目情报服务创造了难得的机遇。

1.1 出版商的书目情报数字化存档

世界上最大的出版学术期刊、会议论文及图书的商业性和学术性出版机构,如 American Chemical Society、AMA、ACM、BCS、IEEE、IEE、IoP、APS、Elsevier、Springer(包括 Kluwer)、Taylor & Francis、John Wiley & Sons、Blackwell、SAGE、Oxford University Press、Cambridge University Press、Emerald Press 等,以及其他很多小型出版商,所提供的开放存取书目记录及文摘不计其数。其中有的机构是直接提供书目情报的开放存取,有的则是通过一些中介机构,如 High Wire Press、MetaPress、IngentaConnect 等。仅 High Wire Press 一家

^{*} 本文系国家社科基金资助项目“数字时代目录学理论的创新与发展研究”(05BTQ010)系列论文之一。

就提供超过350万条开放存取书目记录。以 Elsevier 的 Scirus 服务为例,它共为 Elsevier 出版的1700种期刊中的所有论文提供总数超过650万条的书目记录和文摘的开放存取服务。此外,Scirus 还收录并整合了将近2000万条来自 PubMed 及各类型学术期刊数字化存档(如 IoP, and API)的索引/文摘记录^[4]。

出版商数字化存档对书目信息和文摘数据开放存取仍有一定限制。首先是时间限制,即开放存取的内容被限制在特定的时间段内,比如除最近12个月外,过去10年内的所有出版物的书目信息;其次是功能限制,比如 Ebrary 服务,它整合了来自150多家学术、商业和专业出版机构的3.5万多册权威图书和文献,对于这些图书的书目信息、文摘及内容,Ebrary 虽然提供开放存取和展示,但对所选图书页面的打印和保存仍实行收费。

1.2 开放存取书目数据库

商业性书目数据库提供商也纷纷加入到开放存取的行列,成为重要的获取网络书目情报开放资源。其实在“开放存取”概念提出之前,这种基于开放和共享的书目情报服务就已经出现。

The National Information Services Corporation (NISC)是一家积极从事慈善事业的公司,在“开放存取”出现的十年前,它就开始以“开放存取”方式或以极低的收费向世界上最不发达的国家提供书目情报服务。例如,该公司为所有与之签订协议的用户免费提供 AIDSearch and Child Abuse、Child Welfare and Adoption这两个数据库的即时访问。

著名的 EBSCO,多年来都以开放存取方式向用户提供其索引和文摘数据库——Teachers' Resource Center 和 Library Resource Center。最近,它以更大型的开放存取 Library and Information Science & Technology Abstracts(图书馆及信息科学技术文摘,LISTA)数据库取代了后者。

除了数据库提供商,政府机构和社会团体也是书目数据库的开发者之一;除了生产 Medline、ERIC、AGRICOLA、NTIS 和 Energy Science & Technology 这样著名的索引/文摘数据库,也开发出 ASCE Civil Engineering、PILOTS 这样的开放存取索引/文摘数据库。

INASP(International Network for the Availability of Scientific Publications)是牛津大学的一个慈善组织,由它构思并发起组织了 African Journals Online (AJOL)项目,该项目对推动低收入国家和地区开放存取数据库的发展确立了可供借鉴的模式。AJOL 收

录了来自非洲国家230个期刊中出版论文的书目情报和文摘,共有近2万条论文记录。尼日利亚和南非是其中收录书目数据最多的两个国家,但即便是最不发达的布基纳法索、莱索托、马拉维和苏丹等国家,也都可以通过参与该项目,免费获取书目情报服务和文献传递服务。INASP 在广大不发达国家,通过发起并组织 AJOL 这样的项目,推动当地学术期刊的出版,并创建索引/文摘数据库,提高当地科研成果的可见度,同时也促进了对大型学术出版商全文存档的开放存取^[5]。

2 开放存取资源的目录组织及发现工具

BOAI 对开放存取的实现途径的描述是:自我存档和开放存取期刊。其中,自我存档又可以通过多种方式来实现,即 E-print、作者个人网站、学科数字存档(Disciplinary Archives)、机构数字存档和知识库等^[6]。它们和目前正日益壮大的开放存取期刊一起组成了网络环境下庞大的开放存取学术资源。但是,在目前的开放网络环境下,如何快速便捷地组织、发现和获取这些资源,是关乎开放存取能否得到可持续发展的一个重要方面。开放网络环境下,也需要借鉴和发展目录学的原理和思想,对开放存取学术资源进行科学地揭示和有效地报导,满足人们对它的特定需要,而不必在各个系统间来回奔波。

书目情报工作的开展需要投入大量的时间和精力,来对书目情报进行输入、组织和校验,但是书目情报的编辑工作一旦完成,就能够为用户提供一种组织严密、监管得力的资源发现服务,在开放网络环境下也是如此。

2.1 E-print 资源的组织发现工具

E-print 包括 preprint(预印本)和 postprint(已正式出版文章)两种,都是重要的免费数字化学术资源。E-print 资源十分丰富,据统计,英国在 ac 和 uk 域名下提供 E-print 服务的站点就有21个。E-print 系统彼此互不相干,相关领域或不同领域的资料分散存储,限制了内容的发布与获取的效率,也阻碍了学术群体之间的交流。如果没有组织化的 E-print 资源传播途径,科研人员将自己的成果向外界广泛传播或发现其他科研成果的过程就必须完全依靠自己的力量。

数据库提供商、出版商网络门户和搜索引擎等也都可以作为 E-print 资源书目情报向公众传播的途径。但这些方式都有自己的缺陷,首先数据库商为了

获得运营经费,有些仍实行收费服务,弥补元数据收集和成本;其次,商业出版门户,如 ScinenceDirect 和 SpringerLink 等,通常仅限于对自己的科研成果进行报导和组织;最后,像 Google 这些搜索引擎,虽然拥有精良的数据库和高效的查询算法,但是在学术信息的分析和鉴别方面仍存在很多不足。

近年来,已经陆续有 E-print 开放书目情报服务出现。Scirus 采用了新的搜索技术,可以发现一般搜索引擎无法涉及的来自专业数据库和 E-print 文库的论文,如来自 arXiv.org、Cogprints、NASA、The Chemistry Preprint Server、The Mathematics Preprint Server 等的 E-print 预印本元数据。OALster 是密歇根大学数字图书馆项目,定期对支持 OAI-PMH 的数据库进行元数据收割,对结果进行组织,然后提供服务,截止到 2004 年 12 月,共收录来自 390 家机构的包括预印本在内的 474 万多条书目记录。Citebase Search 是 1999 年的开放引文计划(Open Citation Project, OpCit)的产物,是一个集检索与引文分析为一体的工具,它也能自动搜索 arXiv.org、Cogprints 和 BioMed 的预印本资源,并进行二次开发^[7]。RePEc 是一个经济学领域的 E-print 的元数据采集和组织计划,其中收录了 US Federal Reserve Banks(美国联邦储备银行)、DEGREE(荷兰的一项收藏工作文件的国家项目)、SWoPEc(斯堪的纳维亚半岛的一个国家先导计划)的文档,标识的记录中工作文件已经达到 145000 种,论文 155000 种,软件 1500 种,图书及其章节 750 种。这些元数据馆藏再通过 RePEc 的数据门户提供检索和展示,如 EconPapers(<http://econpapers.repec.org/>)、IDEAS(<http://ideas.repec.org/>)、Inomics(<http://inomics.com/>)和 Socionet(<http://socionet.ru/>)等^[8]。

2.2 开放存取知识库发现和检索工具

开放存取知识库(Open Access Repository)主要有学科知识库(Disciplinary Repository)和机构知识库(Institutional Repository)两种类型,也是重要的自我存档方式。

(1) 开放存取知识库的报导和控制。

首先,开放存取知识库的发现和报导可以通过文档累积列表的方式来实现,但是在学术科研领域很难保证列表的功能和质量。简单的列举无法为用户揭示知识库的信息内涵,以及与信息的利用相关的属性和特征,无法提供结构化信息服务,例如对每个知识库的著录信息进行检索、过滤、分析等。因此,需要为开放存取知识库构造全面而权威的报导和控制途径。

目前通常采用的方式有注册(registry)服务和目录指南服务。

比较著名的开放存取知识库注册项目有:ROAR(Registry of Open Access Repositories,英国南安普敦大学的“开放存取机构知识库注册”),Experimental OAI Registry(美国伊利诺伊斯大学的“实验性 OAI 注册”)。它们都是基于 OAI 协议的注册服务。

这种注册机制包括数据提供方(DP)和服务提供方(SP)的注册、信息发布和协议检测。注册过程中,开放存取知识库作为数据提供方,由管理员在注册服务器列举的学科领域中选择知识库所属的学科(允许多选),并选择知识库提供数据的开放级别,如只开放元数据、开放元数据和部分最终文档、开放元数据和所有最终文档等。如果知识库经检测符合 OAI 协议,注册服务器就把该知识库的有关信息存放在注册数据库中,包括:名称、BaseUrl、采用的 OAI 协议的版本、其他描述信息、支持的元数据格式及相应的 DTD 或 XML schema、注册时间、所属学科领域等。而知识库信息发布则可以通过三种方式来实现:其一,简单列出所有已注册知识库列表,用户可以选择列表中的任意一项来查看相应知识库的详细信息,并可以选择排序方式,如按照名称、元数据数量、数据开放级别等;其二,按一定学科分类标准对已注册知识库进行分类,允许按学科查找;第三,提供检索功能,分为简单检索和高级检索,高级检索允许用户对检索提问进行学科、开放级别等条件的限定^[9]。目前这种注册服务还不是很完善,但随着开放资源的不断增多,这种信息交流平台将会发挥越来越重要的作用。

其次,是知识库目录指南服务。其代表是 OpenDOAR(Directory of Open Access Repositories,开放存取机构知识库目录),由 University of Lund 和 University of Nottingham 共同创建。OpenDOAR 也能够揭示知识库的详细信息并进行分类组织,如同行评议/非同行评议、主题覆盖范围、内容适用群体、收藏和保存策略等。OpenDOAR 与 OAI 注册的不同之处在于,它以一种普通 OAI 注册不支持的方式对知识库加以类别。OAI 注册要求知识库提供所有详细信息,因而有很多开放存取知识库都没有在其中注册。而 OpenDOAR 则通过积极地对知识库进行搜索完成自己的列表,注册程序更为简捷,所收录的知识库的数量和可见度更高。此外,OpenDOAR 还允许用户按照知识库的内容类型和服务群体进行列举,大大提高了查询的准确度^[10]。

(2) 知识库中开放存取资源的发现和检索。

目前,搜索引擎仍然是检索开放存取资源的重要工具。开放存取学科及机构知识库的丰富资源吸引了大批搜索引擎的关注,Google、Yahoo!等主流搜索引擎都希望借用开放存取资源及知识库的内容来增加自身的可用性、访问量和广告收入。其中 Google Scholar 和 Google Print 在这方面走到了前面^[11]。例如可以使用 Google Scholar 来检索学术论文、使用 Google Print 来检索图书。

一定的技术规范也是集成分布式开放知识库资源、提供统一搜索入口的有效途径。2001年1月 OAI (开放存档先导计划)项目推出了 OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting specification),该协议较好地解决了分布式的异构数据库之间的互操作和跨库检索问题。随着 OAI-PMH 在开放存取知识库建设中的应用,开放存取知识库的发展取得了重大突破^[12]。

2.3 开放存取期刊目录

开放存取期刊目录,可将物理分散的信息从逻辑上集中起来,实现分散建设,集中使用。

DOAJ (Directory of Open Access Journals, 开放存取期刊目录)是由瑞典 University of Lund 图书馆创建和维护一个建立在互联网上并通过互联网可免费获取全文的实施质量控制的学术性电子期刊目录。截至2006年6月,该目录收录的开放存取期刊已达2303种,期刊论文已达103870篇,其中有663种期刊可实现论文搜索。共由包括期刊查找、新刊目录、全文搜索、期刊推荐等在内的10个功能模块组成^[13]。

DOAJ 可以提高开放存取学术期刊的可见度、使用率和影响力,为用户提供一站式服务,可以看做是开放存取期刊的网络书目控制和检索工具。

首先,在资源著录和元数据描述方面,DOAJ 采用基于 OAI-PMH 协议下的非限制都柏林核心集。根据描述对象分为两类:一类是期刊元数据,另一类是论文元数据。点击期刊名称就能得到该期刊描述著录内容。

其次,DOAJ 展示的目录类型有刊名目录、新刊目录和主题目录。对于内容涉及多个主题的期刊,则分别在不同的主题类目下重复揭示。

最后,为了实现检索功能,DOAJ 还分别对论文元数据中的6个字段建立索引库,包括题名、期刊名、ISSN、作者、关键词和文摘。通过索引库可以方便获取某一论文的元数据、所属期刊的元数据以及它的全

文链接地址^[14]。

这种以 DOAJ 为代表的开放存取期刊目录,一方面建立期刊推荐平台,扩大开放存取期刊信息的来源;另一方面,方便用户直接、及时地获得期刊所有者提供的期刊和全文元数据。

2.4 开放存取引文索引

2001年,英国 University of Southampton 通过“开放引文计划”(Open Citation Project)创立了 Citebase 服务。Citebase 是一个集检索与引文分析为一体的工具,其特色就在于它可以为开放存取资料提供引文与影响力分析及参考文献链接服务。Citebase 自动搜索 arXiv、Cogprints 和 BioMed 的开放存取文档,并进行二次开发。目前可实现的功能包括:

检索——系统提供了元数据、引文以及 OAI 识别码三种检索途径。在元数据检索界面,用户可以输入作者、题名/文摘关键词、出版物名称、创建日期进行检索,并且可以选择按论文被引数、作者平均被引数、创建日期、最近更新日期、论文点击率、作者平均被点击率等对结果进行升序或降序排列。

引文分析——检索结果可选择以文摘/引文方式展开,除显示来自 OAI 收割的书目元数据和文摘,还包括特色服务,如列出文章被引用和点击历史的表格和曲线图、该文的参考文献列表、引用该文的前5篇文章、引用该文的所有文章、与该文同被引的前5篇文章、与该文同被引的所有文章。这些文章也能以同样的方式展开,理想的情况是通过一篇文章可无限伸展下去,这样能够很好地揭示文献之间的引证关系。

参考文献解析——系统还对文章的每篇参考文献都作了来源解析,即 E-print (arXiv.org 预印库)、journal (刊物)及 G (其他),并作了指向原文的链接,点击 G 则直接调用 Google Scholar 测试版 (<http://scholar.google.com>) 进行检索^[15]。

2005年,英国联合信息系统委员会 (JISC) 公布了报告“Open Access Citation Information”。该项目研究目的是为全球开放存取资源的引文服务确定一个框架,为引文信息的收集和发布、服务需求等创建一种理想化的结构。该项目还为“开放存取引文索引服务”(Open Access Citation Index Service)提出了一个“初步建议”,包括:在机构知识库软件中整合参考文献解析工具,实现对作者上传论文的参考文献数据的自动化提取;自动化解析免费文本中的参考文献格式,并向作者以交互方式反馈析出信息;为开放存取机构知识库建立与被引文献数据库进行交互的标准

化软件,例如通过一个网络服务界面,实现知识库管理者在数据商之间进行选择(例如, CrossRef, PubMed, CiteULike等);创建及调适参考数据库服务,来支持参考文献远程链接,即利用从数据库中自动解析出的数据片断,实现对规范的参考文献的获取、扩展和链接;开发一种标准化的方法,来储存和交换参考文献数据,例如使用 OAI-PMH 的 OpenURL Context Object 标准。

为开放存取资源建立引证关联,可以形成一个学术信息深层挖掘工具和跨学科的开放系统。以上提到的研究均处于试验和探讨阶段,网络开放存取引文索引的利用空间还有待进一步开发。

3 编目及书目数据通用开放标准

在开放网络环境下,开放标准实际上还没有确切的定义。有些人认为只要使用上具有开放性,该标准就可以被称为开放标准;另一些人则认为,不仅是使用环节,标准的创建和调试都以开放程序进行的,才可以称为开放标准。符合前一种定义的有 XHTML 等。而都柏林核心元素集(Dublin Core)则是一种完全的开放标准,它不仅在使用上开放,而且在其运行程序上也是开放的^[16]。

开放标准是开放网络发展的必然产物。首先,开放标准可以在一定程度上克服技术淘汰和硬件依赖问题,保持数字信息的长效获取、长期保存和系统间数据的转移。其次,使用开放标准还有利于实现系统间的互操作,在数字图书馆、图书馆目录及其他系统间“通过互操作实现信息的自由流动”^[17]。

目前,有很多组织都开始倡导开放标准,例如 OAI(Open Archives Institute),提出了基于元数据的电子文献互操作框架,并提出了一项元数据收割协议 OAI-PMH(Protocol for Metadata Harvesting),其目标是通过元数据收割模式实现 Web 上发布信息的不同组织之间的互操作,为它们提供一个与应用无关的互操作框架。该协议规定了两个角色,即服务提供方和数据提供方。数据提供方负责元数据的生成与发布,即为数据库中的数字对象建立元数据并对其进行结构化组织,然后在确保各类资源的相对独立与稳定的前提下,通过元数据的映射,发布符合 OAI 协议规范的元数据(DC);服务提供方通过元数据收获机(Metadata Harvester)从数据提供方和其他服务提供方处收获元数据,并对这些元数据进行加工处理,提供增值服务,建立元数据之间的关系,向用户提供统一的查

询界面。

其他重要的书目数据通用开放标准还有:MARC(Machine-Readable Cataloguing)、MODS(Metadata Object Description Schema)、METS(Metadata Encoding & Transmission Schema)、XOBIS(XML Organic Bibliographic Information Schema)、cistylex 等^[18]。其中 MODS、METS、XOBIS 是书目数据标准;cistylex 是书目格式化标准;而 OAI-PMH、MARC 和 Dublin Core 等都是书目数据检索及链接标准^[19]。

开放网络环境下,书目元数据及其格式、检索及链接标准的开放性是资源获取开放性的重要前提。开放标准正好可以消除元数据互操作和交流上的障碍,也就是消除书目情报服务的鸿沟,实现书目情报资源和书目情报服务公平与开放。

4 网络书目情报服务与开放目录

网上书目情报服务的开展,拓宽了目录学的应用空间。搜索引擎、指引库等秉承了人工书目情报服务原理、方法和程序,是在网络环境下书目情报服务的新发展,是网上智能化的书目情报服务系统^[20]。以搜索引擎为代表的网络书目情报服务,作为社会的重要信息公共平台,不仅影响着每一个社会成员,也需要社会成员的广泛参与;书目情报服务也应该积极地与之合作,扩大社会影响力,提升服务效果。

4.1 开放目录与社会化参与

Open Directory Project(ODP,开放目录项目)是目前互联网上最大、最广泛的人工主题目录。它建立在开放源代码(Open Source)理念的基础之上,成功借鉴了牛津英语词典发动志愿者参与的经验,形成一个由来自世界各地的志愿者共同维护与建设的最大的全球目录社区。截至 2006 年 6 月,已注册来自全球 73354 个志愿编辑者,设置类目多达 59 万多个,共收录站点 400 多万个^[21]。ODP 与其他网络分类目录的主要不同表现在:为网络书目情报服务提供了一种自我组织与管理的模式;百分之百提供免费服务,包括编目数据的下载和使用;类目设置更能覆盖一些边缘学科和冷门学科,性能优于 Yahoo!。目前 ODP 还为互联网上众多搜索引擎和门户网站提供目录服务,对象共包括 Netscape、AOL、Google、Lycos、HotBot、DirectHit 等在内的 332 家网站。

为了更好地满足开放网络环境的动态需求,ODP 在网络书目情报服务创新上表现突出,主要体现在开放目录的编制和管理方式上^[22]。

(1) 志愿者管理规范。包括:①开放申请制度, 不限制志愿者的国别、文化背景、兴趣爱好等;②层级管理制度, 志愿者的申请由高层级的 ODP 编辑批准, 一个编辑员只能有一个账户, 一个月没有登陆记录或连续四个月未进行编辑操作, 账户将被暂停;③编辑规则, 新注册编辑通常只能申请加入某一小类, 上级类别编辑员有权对下级类别编辑人员的编辑处理进行修改, 编辑员可以利用论坛或者电子邮件进行交流;公平编制原则, 对一些网站不负责任地提交或登录、歧视或干预竞争者的行为进行了限制并给出相应的处罚条例。

(2) 编辑操作的规范。包括:①资源选择和评价规范, 要求完整性、独创性、时效性、可用性;②类目设置和处理规范, 具体有子目建置要点和类目调整要点, 为了便于编辑员和网上搜索者了解类目含义, ODP 还规定了类别描述的规范, 其类别描述是开放性的;③标引规范, ODP 对于标引规则的规定总体上较简略, 还采用实例、参考和交流形式作补充说明, 在个别类目的范畴描述里也包括了相应的分类规定;④资源描述规范, 不仅对客观、公正等原则进行了描述, 还具体到对字母大小写、词法、句法以及它们的显示方式的要求, ODP 著录项目有 URL、标题(title)、描述(description)、编辑注释(editor notes)、日期(sort date field)、酷站(cool site feature)。

开放目录的思想不仅存在于网络信息目录组织中, 在传统图书馆目录工作领域中也有所体现。在 OCLC 的 Open WorldCat 计划中, 也新增加了社会化网络工具(social networking tools), 可以实现用户向图书馆收藏的权威的 WorldCat 书目数据添加自己的内容。Open WorldCat 系统在界面上设置了“Details”和“Reviews”两个标签, 通过这两项功能用户可以为查找到的相关目录款目, 加入一些有价值的信息。其工作原理是, 用户注册一个个人账户, 即可以立即提交内容。这个账户可以允许用户在任何时间进入该界面, 登录、创建及修改内容。还提供有操作指南, 监管者提交。用户贡献的内容要与 WorldCat 记录衔接, 但是数据的维护是分别进行的, 并且这些用户新添内容不会成为权威书目数据的一部分。

栏目具体功能如下:

“Details”标签——该标签中, 对于某个特定标题来说, 注册用户可以对目录款目的每个字段分别进行内容和注释的输入, 例如一个作品的主题摘要, 或者历史背景等。类似于 Wikipedia 百科全书这种基于网

络的公共维护的款目, 这些字段信息可以由所有注册用户自由编辑、添加、修订或更新。

“Reviews”标签——在该标签中, 对于某一特定标题, 注册用户可以提交他们自己对作品的评价和观点, 同时可以阅读其他用户的观点。不过这里的内容只有该内容的原始作者才有权修改, 注册用户名将与评价观点一同显示^[23]。

用户贡献内容有助于扩展 OCLC 的编目合作, 将非编目职位的图书馆专业人员和普通用户吸纳进来。他们在内容创建和管理中的共享和参与可以培养对书目情报和目录工作的重视, 并激发对图书馆资源更大的兴趣。这种通过社会化网络工具实现用户贡献内容的做法为开放网络环境下的书目情报服务探索了新的服务形式, 并且为目录工作专业人员提供了一条与用户进行交互、收集用户信息的途径。

4.2 开放目录与整合服务

在目前的网络环境下, 搜索引擎和图书馆目录都可以作为知识入口网站, 它们之间应正视对方在网络书目情报服务中的重要性, 通过合作提供整合服务。正是基于这种理念, OCLC 于 2004 年初, 将 WorldCat 的馆藏数据陆续加入 Google 及 Yahoo! 两大搜索引擎及其工具栏中, 即 Open WorldCat 计划。现在, 最新的 Yahoo! 工具栏、Google 工具栏以及 Firefox 的搜索拓展已经集成了 OCLC 的 WorldCat^[24]。

Open WorldCat 计划使用户可以通过简单的互联网搜索、书目网站或网上书店等渠道获取 OCLC 在 WorldCat 数据库中的馆藏资源。这使 OCLC 的成员馆更容易通过网络被访问, 用户使用搜索引擎就可以很容易地获取所需书目信息。例如, 使用 Yahoo!、Google 等搜索引擎的用户, 当输入一个与图书馆馆藏资源相匹配的条目时, 返回的检索结果将包含指向 WorldCat 的接口信息, 用户还可以输入自己的地理信息, 就可以在他所在的城市、地区或国家最方便的图书馆找到该信息资源。

Open WorldCat 计划将图书馆目录及文献资源向网络搜索引擎开放, 创造了一种开放目录整合服务, 推动信息资源公共获取, 形成了图书馆、用户、搜索引擎和 OCLC 共赢的局面。首先, 对图书馆而言, 通过 Open WorldCat 计划, 创建图书馆频道(library channel), 完成图书馆资源和搜索引擎的无缝链接, 扩大了图书馆馆藏资源的可见度和显示度, 达到吸引更多用户利用图书馆的目的;其次, 对用户而言, 用户可以通过搜索引擎直接检索 WorldCat, 免去了以往必须经

由 FirstSearch 检索的麻烦,完成一站式检索需求;再次,对搜索引擎而言,Google、Yahoo! 通过与 OCLC 的合作,拥有了对 WorldCat 资源的链接权,图书馆频道的创建不仅扩大了搜索引擎的数据库范围,也使得用户对搜索引擎的依赖程度更加提高^[25]。

5 目录强化(bibliographic enrichment)与信息存取

数字时代,用户查询目录不单是为了获取关于资源的效用信息,更重要的是希望通过目录直接获取文献的内容。网络资源发展的现状决定了任何一个机构或系统都无法凭一己之力集中所有的虚拟资源,因而网络虚拟资源的管理可以采用“目录与资源实体分离”的理念^[26],即资源分散建置,目录集中服务,靠目录与资源之间的链接,实现服务的整合和资源的统一获取。

美国国会图书馆(LC)成立了 Bibliographic Enrichment Advisory Team(BEAT,目录强化咨询小组),承担强化书目记录相关研究和实践的任务。该计划主要包含以下内容^[27-28]。

(1)记录的扩展:书目记录与 LC 网络数据的链接。

BEAT 计划正在向 LC 书目记录添加目次(Table of Content, TOC)全文。TOC 原始数据是从图书的目次页扫描的图像中识别信息,现在通过计算机程序将其添加到 505 字段。该字段前端都添加了“Machine generated contents note:”标识。在过去的数年中, BEAT 计划已经从其数字 TOC(dTOC)项目中生成了大约 30000 种图书的网络化 TOC。而机器生成 505 字段的任务就是将这些文档的文本添加到书目记录中去。该项目正在持续开展,它不仅向新的记录添加 TOC,还向其他 BEAT TOC 项目的数据进行回溯添加。

(2)强化书目记录的内容:为研究需要提供更多数据。

添加到书目记录中的数据包括:

文摘和注释——为“A Guide to the Microform Collections in the Humanities and Social Sciences Division”(一个收录大量图书馆馆藏缩微胶片的网络书目)中的每一个分配有 LCCN 号的所有项目的书目记录添加描述性注释。

MARS 最佳注释及评述——即为每年由 ALA's Reference and User Services Association 的 Machine-

Assisted Reference Section(MARS)选出的最佳免费参考资源网站(Best Free Reference Web Sites)的 LC 书目记录添加注释。

参考源评述——在美国图书馆协会的批准下,该项目将 American Libraries 的年度汇编中的“Outstanding reference sources”部分的带有评述的注释添加到相应的 LC 书目记录中。

HLAS 评述——从分散的图书馆 Handbook of Latin American Studies(HLAS)数据库中析出有关专著

的评述,添加到 LC 的书目记录中。

此外还包括向 E-CIP(电子在版编目)记录中添加 TOC 及机器生成的 TOC。

(3)加强存取:书目与电子资源相链接。

连续出版物的网络存取——该项目有几个方面的任务组成:将许多“working paper/discussion paper”的连续出版物与它们的网络电子版相链接;为大量具有较高学术价值的、专业性的、但未在 LC 目录中得以展示的连续出版物创建电子连续出版物记录;为挑选出的公共领域内的可以以数字化形式获得的连续出版物中每篇专论编制书目记录,目前已经为 30000 多篇专论提供了全文电子文本的存取访问。此外,该项目还为所涵盖的连续出版物创建了一个网络数据库——Technical Reports and Working Papers in Business and Economics,网址为:<http://www.loc.gov/rr/business/techreps/techrpshome.php>。

公共领域作品的网络存取——BEAT 的这项行动将 LC 中收录的书目记录与其相应的电子全文之间建立了链接,包括本馆和其他机构的馆藏资源。LC 希望为用户提供更加统一的、集中的资源访问,并提供丰富的、实质性的有关资源的效用信息,以及对全文的存取访问。其中有两个合作伙伴及其近来的添加工作很值得注意:RAND Corporation(兰德公司)和马里兰大学的 Thurgood Marshall Law Library。前者通过 RAND 网站 <<http://www.rand.org/publications>>, 添加了超过 680 个指向 RAND 出版物的链接;而后者,则有超过 140 条链接指向由马里兰大学 Thurgood Marshall Law Library 数字化生成的 United States Commission on Civil Rights 出版物,网址为 <<http://www.law.umaryland.edu/marshall/usccr/index.asp>>。

BeCites⁺——BeCites⁺(Bibliographies plus: Enhanced Citations with Indexes, Tables of contents, Electronic resources and Sources cited,书目添加:强化的引

文索引、目次、电子资源及引用来源)不仅通过为目次、索引和引用来源各自创建及添加链接来强化人工生成的书目,而且还为较早的出版物创建了强化的电子“webliographies(网络书目)”,这些书目都新增加了与该主题相关的电子资源参考文献,作为原始出版物所含内容资料的补充。

6 开放源代码书目软件(Open source bibliographic software)

开放源代码软件(open source software)简称开源软件,是一种源代码公开的软件。这种“公开”意味着用户可以自由地使用、复制、散发以及修改源代码(即补充漏洞,按具体要求定制功能)。近年来,一场以开放源代码软件为代表的,促进知识共享和技术创新为目的的开发模式和变革运动正以势不可当之势在全球兴起。伴随着数字资源建设和目录组织日益在开放分布的网络环境下展开,很多书目工具也逐渐采用开放源代码软件的模式开发和应用。

2004年,美国Indiana University发布了一个开放源代码书目软件——PubsOnline,可用于管理在线书目,有助于通过网络存储和现实引文数据。PubsOnline具有较好的灵活性,它可以用来管理某个基金资助项目成果的被引情况,或者某个机构或多个机构的所有科研经费产出的情况。PubsOnline提供了多项特色功能,比如,它可以完善书目情报,通过定制属性对引文信息进行分类,并向数据库(MySQL)提交书目情报;提供一种直接简明的基于网络的服务方式,便于被授权的个人用户向数据库添加新的款目;简化了网络信息的检索,无须特殊的软件就可以实现信息存取访问;PubsOnline使用与BibTeX相兼容的格式存储信息,用户既可以通过标准HTML页面看到引文信息,也可以BibTeX格式获取引文数据的文本文档。PubsOnline是一个高效的开放源代码工具,是机构组织跟踪和揭示其智力成果影响力的辅助软件^[29]。

PubsOnline只是众多开放源代码书目软件中的一例,其他使用较为广泛的书目软件还有BibDesk、Pybliographer、zNote、OpenBiblio等^[30]。开放源代码软件对网络书目的影响不仅在于提供了经济易用,且具有良好可扩展性的书目数据组织、管理和发布工具,还在于为网络开放目录(open directoly)的组织和管理思想上给予了启发。开放目录正是借鉴了开放源代码软件高度开放、自由修改、合作开发、成本低廉的特性^[31],为网络书目情报服务开创了新的模式。

参考文献

- 1 王云才. 国内外“开放存取”研究综述. 图书情报知识, 2005(12)
- 2 “Budapest Open Access Initiative” ARL Bimonthly Report 220. [2006-07-11]. <http://www.arl.org/news/220/boai.html>
- 3, 4, 5 Peter Jacso. Open Access Scholarly Databases - A Bird's Eye View of the Landscape. [2006-07-07]. <http://www.ifla.org/IV/ifla72/papers/157-Jacso-en.pdf>
- 6 Bailey, Charles W., Jr. Key Open Access Concepts. [2006-07-07]. <http://www.escholarlypub.com/oab/keyoaconcepts.htm>
- 7, 15 冯艳花. 基于OAI的电子预印本资源共享. 情报理论与实践, 2005(4)
- 8 Thomas Krichel. The Economics of Open Bibliographic Data Provision. [2006-07-07]. <http://www.econ.uconn.edu/working/2005-01.pdf>
- 9 朱遵学等. OAI注册服务器功能的探讨. 图书馆杂志, 2004(8)
- 10 About OpenDOAR. [2006-07-09]. <http://www.opendoar.org/about.htm>
- 11 Open access in 2004. [2006-07-09]. <http://www.earlham.edu/~peters/fos/newsletter/01-02-05.htm>
- 12 李武. 开放存取出版的两种主要实现途径. 大学图书馆学报, 2005(4)
- 13 Directory of Open Access Journals. [2006-07-07]. <http://www.doaj.org/>
- 14 谭从容, 阮春英. 关于DOAJ的概况、使用和影响分析. 图书馆建设, 2006(1)
- 16, 18 Edward M. Corrado. The Importance of Open Access, Open Source, and Open Standards for Libraries. [2006-07-09]. <http://www.isrl.org/05-spring/article2.html>
- 17 The Open Group. Developer Declaration of Independence. [2006-07-10]. <http://www.opengroup.org/declaration/declaration.htm>
- 19, 30 Open standards and software for bibliographies and cataloging. [2006-07-10]. <http://www.search.sourceforge.net/bib/openbib.html>
- 20 司莉, 彭斐章, 贺剑峰. 网络信息资源组织与目录学的创新和发展. 图书情报工作, 2001(9)
- 21 Open Directory Project. [2006-07-07]. <http://dmoz.org>
- 22 金海燕, 马张华. Open Directory剖析. 中国索引, 2005(3)
- 23 Additional features: User-contributed content via social networking tools. [2006-07-06]. <http://www.oclc.org/asiapacific/zhen/worldcat/open/features/default.html>

(转第30页)

只有在联盟的整体收益大于单个成员馆投入总和,并均衡所有相关利益之后,联盟的运行模式才能成功。

合理的利益分配机制是建立在成本核算基础上的。在联盟的运行过程中,需要对网络条件下图书馆资源共享的成本核算及补偿原则进行分析和研究。如文献资源成本、系统硬件成本、软件成本、网络通讯服务成本、设备更新及维护成本、培训及教育成本和管理成本等的构成比例和消耗特点,不同的成员馆会有不同的测算方式。联盟应制定统一的成本核算策略和方法,建立自己的成本分配模型,在成本效益分析的基础上对成员馆的贡献率进行分级和界定,对某些成员馆的成本支出可以考虑吸取外来资金、向获益者收取费用以及转向使用图书馆现有资金等方式进行补偿。

4.4 依法规范联盟的权利和义务,建立动态竞争机制

签订联盟协议,是利用法律手段保护联盟成员正当权益不可缺少的部分。在联盟组建之初,应建立大家共同认可的联盟协议,敦促联盟成员遵守约定的游戏规则,同甘共苦,相互监督。同时随着外界条件的变化,比如国家对文化、信息产业政策的调整,图书馆或企业的技术进步,信息资源提供商营销策略的改变,信息服务市场的变化,渐进地修改合作协议的条款,并考虑各成员馆的承受能力,以免引起冲突,导致联盟解体。对于某个联盟成员损人利己的行为,要及

时采取对策,或按协议规定进行仲裁,或终止合作,并要求承担损失。对于联盟协议未作规定而联盟过程中出现的新问题,联盟各方要进行谈判,协商解决。

同时,联盟也应适应市场机制,在联盟内部建立动态的竞争机制,优胜劣汰,惟此才能实现联盟的高效运作和共存价值。联盟各成员馆之间是一种既合作又竞争的关系,联盟应注意协调各方面的利益,把联盟风险排除在建立之前。

参考文献

- 1 [2006-05-01]. <http://www.library.yale.edu/consortia/icolemembers.html>
- 2,3,4 Gary E. Gorman, Rowena Cullen. Models and opportunities for library co-operation in the Asian region. *Library Management*, 2000, vol. 21 No. 7
- 5 [2006-06-05]. <http://www.oclc.org/default.htm>
- 6 罗琳,晏凌. 电子参考咨询的功能成本分析——以武汉大学图书馆为例. *图书馆论坛*, 1999(2)
- 7 Shoaf, Eric C. The effects of consortia membership on library planning and budgeting. *Library Administration & Management* Vol. 13 No4 (Fall 1999)

谢春枝 武汉大学图书馆副研究馆员。通信地址:武汉大学。邮编 430072。(来稿时间:2006-05-24)

(接第24页)

- 24 Open WorldCat-enabled Web tools. [2006-07-06]. <http://www.oclc.org/worldcat/open/searchtools/default.htm>
- 25 张秋. OCLC的OPEN WorldCat计划及其对图书馆信息服务的启示. *图书馆杂志*, 2005(11)
- 26 方威明,白少辉. 基于OAI的虚拟资源联合目录之设计. [2006-07-06]. <http://preprint.nstl.gov.cn/newprint/Upload/2004/1100420094410.pdf>
- 27 BEAT Report for ALA June 2005. [2006-07-08]. http://www.loc.gov/catdir/beat/beat_reports/BEAT_ALA_report_june_2005.html
- 28 The Bibliographic Enrichment Advisory Team (BEAT). [2006-07-08]. <http://www.loc.gov/catdir/beat>
- 29 PubsOnline: Open Source Bibliography Database. [2006-07-

06]. <http://delivery.acm.org/10.1145/1100000/1099492/p247-myron.pdf?key1=1099492&key2=1158142511&coll=ACM&dl=GUIDE&CFID=15151515&CFTOKEN=6184618>

- 31 易晓春. 开放源代码软件及其数字图书馆. *图书馆杂志*, 2003(11)

彭斐章 武汉大学人文社会科学资深教授,博士生导师。通信地址:武汉。邮编 430072。

邹瑾 武汉大学信息管理学院博士生。通信地址同上。

(来稿时间:2006-07-18)