

# 基于元数据仓储的数字资源管理探讨

麦林陈和

(厦门大学图书馆, 福建 厦门 361005)

**摘要:**以厦门大学图书馆为例, 针对目前数字资源存在系统多样性、利用率较低、不易长期保存和不稳定等原因, 厦门大学图书馆为数字资源建立了统一元数据仓储, 并作为馆藏内容一部分。同时, 厦门大学图书馆还在元数据仓储的基础上建立了统一检索应用平台, 实现了数字资源统一检索, 提高了数字资源的利用率和信息服务质量。

**关键词:**数字资源; 元数据仓储; 统一检索; 厦门大学图书馆

中图分类号: G25

文献标识码: A

文章编号: 1672-6251(2011)08-0030-04

## Study on Digital Resources Management with Metadata Repository

MAI Lin, CHEN He

(Xiamen University Library, Fujian Xiamen 361005)

**Abstract:** Taking the Xiamen University Library as an example for the digital resources application systems' diversity, low utilization rate, difficulty of long-term preservation and instability, Xiamen University Library has created a unified digital resources metadata repository as part of the collection. At the same time, Xiamen University Library has established a unified search applications platform based on the repository to improve the digital resources utilization rate and information service quality.

**Key words:** digital resources; metadata repository; union search; Xiamen university library

### 1 引言

数字资源, 亦称电子资源, 是文献信息的表现形式之一, 是将计算机技术、通信技术及多媒体技术相互融合而形成的以数字形式发布、存取、利用的信息资源总和。商业化的数据库、机构或个人建立的数据库、各种网络免费资源等都属于数字资源<sup>[1]</sup>。

随着数字资源在现代图书馆馆藏与服务中的地位不断上升, 以及读者对信息需求方式的转变, 现代图书馆对数字资源馆藏建设与服务方面面临着新的机遇与挑战。

### 2 数字资源馆藏与服务现状

随着信息技术与网络技术发展, 信息资源逐渐电子化和网络化, 图书馆的馆藏资源结构也因之发生重大的变化。20 世纪 90 年代以来, 国内图书馆开始引进各种类型的数字资源, 图书馆馆藏也由单一的传统印刷型资源向传统印刷型资源与数字资源并举的复合馆藏转变, 并有不断扩大数字资源馆藏的趋势。以购置经费为例, 不但购买数字资源的经费逐年增加, 而且占文献购置总经费的比例也是逐年增加, 比如四

川大学图书馆, 该馆 2004 年、2006 年、2008 年数字资源经费占总经费的比例分别是 13.14%、25.09% 和 39.49%<sup>[2]</sup>。

数字资源相比较于传统印刷型资源, 具有使用灵活、不受时空限制、检索速度快、传递速度快、支持多用户同时使用等诸多优点。对于图书馆而言, 采购数字资源减少了验收、编目、典藏、阅览、流通等各环节的人力、物力以及时间和空间的成本。因此, 数字资源出现以来, 一直备受读者用户和图书馆的青睐, 调查显示, 有 90% 的读者使用数字资源, 而教师使用数字资源的比例更是高达 96.2%<sup>[3]</sup>。

随着数字资源种类和数量的增加, 一方面给读者用户带来全新的丰富的信息资源, 图书馆的馆藏种类和数量也迅速膨胀; 另一方面, 种类繁多和数量庞大的数字资源给读者用户带来了使用上的困惑。图书馆面临如何更好地组织管理这些数字资源和如何提供更好服务的问题。具体来说, 目前数字资源馆藏和服务主要存在如下三个方面的问题:

(1) 资源孤岛, 自成系统。全球数字资源提供商

作者简介 麦林(1976-) 男 本科 馆员 研究方向 资源建设、学科服务。  
收稿日期 2011-07-07

成百上千家,再加上数字资源类型不一样,决定了数字资源的应用系统、使用方法、数据格式等方面的多样性。数据商之间的竞争关系,更导致了各个应用系统之间的互不兼容,各自形成封闭的资源孤岛。读者想要获得全面而准确的学术信息往往需要依次进入各个数字资源的搜索界面逐一搜索,并且要对各个数字资源的搜索规则有足够的认识方可获得所需信息。让读者感到困惑的一般不是进入到某个具体数字资源的使用,而是面对数量众多的数字资源无从选择,就像进入一个摆放无序的大超市,虽然商品琳琅满目,查找却很费劲<sup>[4]</sup>。

(2) 数字资源利用率比较低。虽然数字资源越来越受到读者用户的青睐,但是通过调查发现,不同学校 and 不同学历人群之间,对不同类型数字资源的使用频度并不一样,总体上,数字资源的利用率还是比较低的。造成这种情况的原因主要是用户对数字资源的不了解和数字资源使用障碍<sup>[3-5]</sup>。

(3) 数字资源的不易长期保存与不稳定性。在市场经济环境下,每个数据商都有一个生命周期,图书馆的馆藏建设随着学校的学科建设调整也在不断调整,以及技术、经济和政治等因素都可能影响到数字资源的持续服务、维护和升级。这些因素成为数字资源的不易长期保存与不稳定性。

对于本地镜像数字资源,图书馆拥有对其控制权,即使数据商破产倒闭或停止购买或其他原因导致与数据商失去联系,图书馆仍然可以据此提供信息服务。

对于远程数字资源,因其主要保存在全球少量的数据商或集成商的服务器上,图书馆没有对这些资源的控制权,一旦某种原因导致与数据商或数据提供者失去联系,那么这些数字资源的访问与获取也将停止,图书馆对该资源将“片甲不留”,前期对该数字资源的经费投入成了“竹篮打水一场空”,这个结局对所有图书馆来说都是不愿意接受的<sup>[2]</sup>。

### 3 数字资源馆藏与服务创新模式

纵观国内各个图书馆数字资源的馆藏与服务模式,基本上停留在对不同数字资源的简单罗列,读者用户首先根据自己的判断选择不同的数字资源,然后分别进行访问检索。对于学科比较齐全、数字资源比较丰富的图书馆,还提供了学科导航服务,其本质也只是对这些数字资源进行了比较高级别的资源罗列。厦门大学图书馆自20世纪90年代以来,经过多年的数字资源馆藏建设,形成了学科比较齐全、覆盖

学术领域比较广的数字资源馆藏。厦门大学图书馆的数字资源馆藏与服务也同其他图书馆一样,面临着前文所述的问题与不足。为此,厦门大学图书馆以改变数字资源馆藏模式、节省读者时间、提高资源利用率为目标,对本馆的数字资源馆藏与服务模式进行了有益的探索与尝试。

#### 3.1 建立元数据仓储

厦门大学元数据仓储主要是用来收集、存储、管理和维护数字资源元数据的数据库系统,其建设的最终愿景是收集厦门大学图书馆所拥有的所有数字资源的元数据及其他符合本馆馆藏要求的有较高学术价值的免费数字资源元数据。

##### 3.1.1 确立元数据标准

目前,厦门大学图书馆所拥有的数字资源,包括本馆自建的数字资源和采购引进的数字资源,种类繁多,所采用的元数据标准各异。为了使厦门大学元数据仓储有一个统一的元数据标准,同时又兼顾不同的数字资源,经过考察和讨论,采用现在国际比较通用和流行的都柏林核心元数据(Dublin Core, DC)作为厦门大学元数据仓储的元数据标准。各个数字资源的元数据尽射成DC量映元数据。

##### 3.1.2 获取元数据方法

每种数字资源或数据库都是一个单独的系统,而且数据来源也不一样,因此,获取数字资源元数据的途径也将多种多样。归纳起来有如下几种途径。

(1) 元数据收割。通过OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting)收割数字资源的元数据是厦门大学元数据仓储获取元数据的最主要途径。

对于馆内自建的数字资源,如OPAC数据库、厦门大学学位论文库、信息参考库、厦大文库等,由于这些数字资源由图书馆自己控制,可以通过编程的方式为这些数字资源添加OAI接口。元数据仓储通过此接口定期批量收割其元数据。

对于馆内本地镜像资源并且开放底层数据接口的,比如超星电子图书、Apabi电子图书等,获取元数据的方式同馆内自建数字资源。对于馆内本地镜像资源但不开放底层数据接口的,比如中国期刊网、万方数据网等,处理方式同远程数字资源。

对于远程数字资源可以分为两种,一种免费远程数字资源;另一种是商业远程数字资源。对于前者,多数已经开放OAI接口,元数据仓储可以直接收割其元数据,如HighWire、Nature、DOAJ等。对于后者,

有些有 OAI 接口,但不对外开放,有些根本就没有 OAI 接口,所以需要通过和数据商进行谈判协商等灵活方式获取。目前,厦门大学图书馆已成功地与德国科技出版集团 Springer 达成协议,根据协议, Springer 免费为厦门大学图书馆开放其数据库 OAI 接口,厦门大学元数据仓储定期通过此接口收割 Springer 的电子期刊、电子图书和电子图书章节的元数据。

(2) 元数据导出导入。对于没有提供 OAI 数据接口,不能通过元数据收割方式获取元数据的数字资源,通过谈判协商等灵活方式让数据商定期或不定期导出元数据,经过规范处理后再导入到厦门大学元数据仓储中,如 MyiLibrary 电子书。

(3) 元数据购买。对某些数据商明确表示需要额外经费购买元数据的,图书馆可以申请一些经费酌情选择购买需要的元数据,经过规范处理后再导入到厦门大学元数据仓储中。

(4) 网页抓取。对于某些不提供 OAI 接口,也不另外提供元数据,但可以通过 Web 网页访问其内容的数字资源,可以通过网络信息技术手段来分析网页并获取元数据。

(5) 其他。与其他兄弟院校互通有无,交换你有我无的元数据。

### 3.1.3 建立元数据仓储的意义

(1) 作为馆藏建设的一部分。厦门大学元数据仓储存储了各种数字资源除数字对象(全文)外的其他必要信息,可以作为数字资源的参考数据库(书目、文摘、索引)使用,而且是由图书馆负责维护和管理,可以纳入图书馆数字资源永久馆藏的一部分。因此,厦门大学元数据仓储在一定程度上弥补了“数字资源的不易长期保存与不稳定性”的不足。

(2) 为信息增值服务作好数据准备。现代信息服务离不开信息技术的支持,但是信息技术无论如何发展,都是以大量事实数据为基础。因此,在当前建立好各个数字资源的元数据库,将为今后的信息增值服务应用,比如统一检索、关联数据应用、数据挖掘等提供数据准备。

(3) 数据备用功能。厦门大学元数据仓储作为数字资源的参考数据库,可以在数字资源由于网络中断、服务器故障等原因而不能访问时,替代原数字资源提供查询服务和索引功能。

(4) 查重功能。理想状态下,厦门大学元数据仓储囊括了厦门大学图书馆所拥有的所有数字资源的元数据,在将来新采购或引进数字资源时,可以对照此

元数据仓储进行查重,避免重复采购,节省采购经费以及人力、物力的消耗。

### 3.2 建立学术搜索

据杨毅、邵敏等学者的调查,有 91.8% 的受访读者是通过图书馆馆藏资源获取学术信息,有 78.0% 受访读者是通过搜索引擎获取学术信息<sup>[3]</sup>。而国外一份调查也显示相似结果,96% 受访读者是通过图书馆馆藏资源获取学术信息,92% 受访读者是通过搜索引擎(如 Google)获取学术信息<sup>[6]</sup>。这些调查结果显示,搜索引擎已经成为读者用户获取学术信息仅次于图书馆的主要渠道。通过分析,之所以读者用户喜欢使用搜索引擎,是因为搜索引擎具有界面简洁、使用简单、海量收录、快速响应等优点。

厦门大学图书馆根据用户获取学术信息行为习惯和借鉴搜索引擎的特点,在厦门大学元数据仓储的基础上,开发建立了数字资源的统一检索应用平台——厦门大学学术搜索引擎。厦门大学学术搜索引擎将有如下功能和意义:

(1) 实现数字资源统一检索。厦门大学学术搜索是在拥有厦门大学数字资源元数据的元数据仓储基础上建立的一个搜索引擎,可以实现厦门大学不同数字资源之间跨库检索。用户不需要专门培训,也不需要专门学习使用方法,与谷歌(Google)、百度(Baidu)搜索引擎类似,只要输入自然语言就可以实现在不同数字资源之间同时快速检索,检索结果以统一格式显示。

厦门大学学术搜索实现了不同数字资源之间的信息整合与应用,在一定程度上克服了数字资源“资源孤岛,自成系统”的不足,节省了读者用户学习数字资源使用方法和检索获取信息的时间。

(2) 提高数字资源利用率。厦门大学学术搜索平台为数字资源的揭示和利用增加了一个便利渠道,必将增加数字资源的利用率,从而更大程度上发挥了投入到数字资源建设中巨额经费的服务效益。

(3) 扩大数字资源的知名度。在数字资源“各自为政”的状态下,由于“马太效应”的作用,利用率愈高的数字资源被用得愈多,而利用率愈低但质量更好却不知名的数字资源愈被用户冷落。在厦门大学学术搜索平台下,各个数字资源被检索、被利用的机会是均等的,只有质量好的数字资源利用率才会高,对于不知名的数字资源可以据此提高知名度。

(4) 减少培训成本。图书馆在购进或引用新的数字资源后,每学期都要花费大量的人力和物力举办数

字资源使用培训班,或者派专门人员到各个院系进行宣讲,培训、宣讲的主要内容是教授用户如何使用不同的数字资源。厦门大学学术搜索屏蔽了不同数字资源使用方法的差异,为不同数字资源使用提供了一个便捷渠道。因此,图书馆在进行数字资源培训、宣讲时,可以借厦门大学学术搜索略去一些数字资源的培训、宣讲。

#### 4 问题与讨论

厦门大学图书馆对数字资源馆藏与服务模式进行了初步探索,建立了厦门大学元数据仓储和统一检索应用平台,在这过程中遇上了如下问题,还有待进一步探索和实践。

##### 4.1 元数据获取

能获取多少量的元数据是厦门大学元数据仓储建设成功的关键。目前,厦门大学元数据仓储可以成功获取本地自建数字资源、开放数据接口的本地商业数字资源、多数免费的远程数字资源的元数据,而如何获取远程商业数字资源元数据则是厦门大学元数据仓储建设遇到的最大问题。

远程商业数字资源非图书馆所能控制或管理,要获取其元数据必须与数据商进行协商谈判。厦门大学图书馆目前暂时还未完成与全部数字资源数据商关于获取元数据的谈判,但从沟通过的几家数据商来看,问题主要集中在付费与免费的问题上,图书馆希望不需要另外付费就能够免费获取数字资源的元数据。某些数据商本着友好合作的态度,很乐意免费提供其元数据,比如本案与德国科技出版集团 Springer 合作就是一例成功的案例。

因此,建议图书馆资源采购部门在采购新的数字资源或续订已有的数字资源时,增加谈判内容,把类似“数据商有义务免费提供图书馆已采购资源的元数据给图书馆永久二次使用”的条款添加至采购协议或续订协议中。

##### 4.2 产权风险

厦门大学元数据仓储并不拥有和存储数字资源数字对象本身,只是存储数字资源的元数据,并在此基

础上提供信息增值服务。这些存储在本地的元数据是否侵害了数据商或其他版权人的利益?从现有文献来看,暂时还没有相关侵权案例的报道。

##### 4.3 技术问题

厦门大学元数据仓储元数据条目数已经达到千万条,属于海量数据级别,随着时间的推移,数据量还会迅速增加。如何存储、维护和管理这些不断增长的庞大的电子数据对于图书馆来说是一项“史无前例”的任务,将面临不小的技术挑战。而如何更好地利用这些丰富的二次文献数据,比如统一检索、关联数据、资源共享等,将有待进一步探索和实践。

#### 5 结束语

数字资源在当今乃至今后的信息社会里必将发挥着越来越重要的作用,图书馆将不可避免遇到如何利用数字资源优化馆藏并提供优质的信息服务的问题。充分利用现有的数据资源、现代信息技术和先进理念进行数据挖掘是一条路子,本文希望在这方面能起到抛砖引玉的作用,从而引起更多同仁对数字资源馆藏与服务的探讨和实践。

#### 参考文献

- [1] 周明华.电子/数字资源馆藏发展政策的制定[EB/OL].2009-5.
- [2] 李咏梅,袁学良.论数字资源与纸本资源的协调发展[J].中国图书馆学报,2009,35(182):51-57.
- [3] 杨毅,邵敏,等.数字资源建设与利用的读者调查——由读者调查结果分析读者利用数字资源的方式与倾向[J].大学图书馆学报,2006,(6):39-48,60.
- [4] 丁枝秀.普通本科高校图书馆数字资源利用情况调查分析——以南京信息工程大学图书馆为例[J].农业图书情报学刊,2009,21(6):19-22.
- [5] 勾丹,郑洋洋.我国高校图书馆数字资源利用现状与对策[J].图书馆学研究,2008,(11):85-87.
- [6] Alison J. Head, Michael B. Eisenberg. Truth Be Told: How College Students Evaluate and Use Information in the Digital Age.[2011-03-15]. [http://projectinfolit.org/pdfs/PIL\\_Fall2010\\_Survey\\_Full-Report1.pdf](http://projectinfolit.org/pdfs/PIL_Fall2010_Survey_Full-Report1.pdf).