



图书馆馆藏数字资源 开发与管理

韩新月 陈月婷
许玉梅 李 丹 著

国家图书馆出版社

图书馆馆藏数字资源开发与管理

韩新月 陈月婷 许玉梅 李 丹 著

国家图书馆出版社

图书在版编目(CIP)数据

图书馆馆藏数字资源开发与管理/韩新月等著.--北京:国家图书馆出版社,2017.12

ISBN 978-7-5013-6187-8

I. ①图… II. ①韩… III. ①数字图书馆—信息管理—研究
IV. ①G250.7

中国版本图书馆 CIP 数据核字(2017)第 194220 号

书 名 图书馆馆藏数字资源开发与管理
著 者 韩新月 陈月婷 许玉梅 李 丹 著
责任编辑 金丽萍

出 版 国家图书馆出版社(100034 北京市西城区文津街 7 号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706(传真) 66126156(门市部)

E-mail nlcpress@nlc.cn(邮购)

Website www.nlcpress.com →投稿中心

经 销 新华书店

印 装 北京鲁汇荣彩印刷有限公司

版 次 2017 年 12 月第 1 版 2017 年 12 月第 1 次印刷

开 本 880 毫米×1230 毫米 1/32

印 张 6.375

字 数 160千字

书 号 ISBN 978-7-5013-6187-8

定 价 48.00 元

前言

图书馆承担着文化保存与传承的社会职能,蕴藏着丰富的信息资源。随着文化事业的繁荣发展,在国家法律政策、社会公众的文化产品需求以及现代信息技术的共同作用下,图书馆的馆藏资源开发与利用活动日渐活跃,在一定意义上成为图书馆积累资源、发展服务的新增长点。

图书馆进行馆藏开发的形式多种多样,从基本的编目整理,到实体馆藏资源的数字化加工,直至发展成功能丰富的数据库系统以及文化创意产品,无不体现了图书馆深入挖掘馆藏资源价值与深化服务的目标导向。在经济社会发展的新形势下,围绕馆藏数字资源开展的开发与管理活动是图书馆进行转型发展模式探索的重要方面。因此,有必要从理论和实践层面研究图书馆馆藏数字资源开发与管理的相关问题,研究开发与管理活动的内在规律,探索科学的开发管理方法,进而提高馆藏数字资源开发活动的整体水平,实现图书馆价值提升,推进图书馆事业与文化科学事业的健康有序发展。

本书运用理论与实践相结合的研究方法,在梳理馆藏数字资源开发与管理的范围与环境的基础上,以工作流程为主线,突出开发规划、数字资源加工、组织与整合等重点环节,分为六个部分:①界定馆藏数字资源的范围,探讨馆藏数字资源开发与管理的內容,分析其影响因素,并梳理常见的开发方式与一般开发流程;②研究馆藏数字资源的开发环境,包括经济社会环境、法律政策环境以及信息技术环境,为开发活动的规划与实践提供支持;③探讨馆藏数字资源开发与管理的规划问题,了解规划功能,明确规划原则,总结分析制定规划的常用方法,着重建立馆藏数字资源开发规划的一般流程;④落实馆藏数字资源加工实践问题,探讨数字资源加工技术、标准规范以及长期保存的理论与实践问题,并分析加工实例;⑤研究馆藏数字资源的组织与整

合,梳理适用于馆藏数字资源开发的资源组织方法,分析资源整合思路与整合方式,同时,也介绍了若干资源揭示与整合的实例;⑥着眼于馆藏数字资源开发与管理活动的宏观层面,研究开发项目的运行与管理问题,分析开发项目的组织方式以及各个主要环节的管理要点。此外,本书也对图书馆馆藏数字资源开发与管理活动的未来发展做了一定的前瞻性分析,以期对图书馆当前以及未来的实践工作有所裨益。

本书作者来自图书馆工作一线,参与了多项馆藏数字资源开发与管理的工作。本书的内容建立在研究团队共同讨论与合作研究的基础上,韩新月负责统稿、校对以及修订和增补相关内容,并独立完成第2章、第3章及结语部分的撰写;李丹负责撰写第1章、第6章,陈月婷负责撰写第4章,许玉梅负责撰写第5章。特别感谢本项目发起人李晓明主任对本书研究思路与结构框架的定位。国家图书馆出版社对本书出版给予了帮助和支持,在此致以诚挚的谢意!此外,本书的写作参考和引用了大量专家学者、业界同仁的研究成果,在此一并致谢。

由于经济社会环境与文化事业不断发展变化,与图书馆馆藏数字资源开发与管理密切相关的理论、方法与技术都处于快速发展之中,加之作者本身学识水平有限,本书的研究内容难免有所疏漏和谬误,敬请专家学者和读者批评指正。

作者

2017年5月

目 录

前 言	(1)
1 概述	(1)
1.1 馆藏数字资源的含义	(1)
1.2 馆藏数字资源开发与管理意义	(6)
1.3 馆藏数字资源开发与管理的内容	(8)
1.4 馆藏数字资源开发与管理的影响因素	(9)
1.5 馆藏数字资源开发方式与流程	(12)
2 馆藏数字资源开发环境分析	(17)
2.1 经济社会环境	(17)
2.2 国家法律政策及行业导向环境	(20)
2.3 信息技术环境	(39)
2.4 图书馆微观环境	(43)
3 馆藏数字资源开发与管理规划	(48)
3.1 战略规划与项目规划	(48)
3.2 规划的主要功能	(52)
3.3 制定规划的原则	(54)
3.4 制定规划的常用方法	(57)
3.5 制定规划的一般流程	(62)
4 馆藏数字资源加工	(74)
4.1 资源加工	(74)
4.2 数字资源加工的技术与标准	(124)
4.3 数字资源保存	(136)
5 馆藏数字资源的组织与整合	(142)
5.1 数字资源组织方法	(142)

5.2	数字资源整合与揭示	(149)
6	馆藏数字资源开发项目的运行与管理	(180)
6.1	项目的组织方式	(180)
6.2	项目运行与管理	(184)
6.3	其他相关事宜	(189)
	结语	(191)
	参考文献	(195)

1 概述

1.1 馆藏数字资源的含义

1.1.1 馆藏数字资源概念

随着信息技术的发展,除了常见的以纸质为载体的信息资源外,以计算机、网络为载体的数字化信息资源迅速增长。关于资源类型描述和界定也不一而足,例如,“电子资源”(electronic resources)、“网络信息资源”(networked information resources)、“虚拟资源”(virtual resources)、“联机信息资源”(online information resources)、“数字资源”(digital resources)等,均是在理论研究和实践中常见的概念,其中,“数字资源”是我国图书馆界应用广泛的一个专业术语,关于其概念定义存在多种观点。一般认为,数字资源是指整个生命周期中要不断地结合计算机媒介形式的信息资源,它们是以数字化形式记录的,以多种媒体形式表达的,分布式存储在非纸质的光、电、磁等载体上的,并通过网络通信、计算机终端的方式传递和再现的信息资源的集合^①。数字资源与传统资源根本的不同在于存储介质,进而导致传播途径和读取方式的不同。

对于图书馆而言,馆藏数字资源是指图书馆馆藏中所有的数字资源,包括图书馆本地拥有的和获得一定期限使用权的数据库和数字文献^②。图书馆可通过购买、许可授权、受缴、捐赠、交换、数字化、网络信

① 毕强,陈晓美等. 数字资源建设与管理[M]. 北京:科学出版社:2010:5—6.

② 吕淑萍,罗云川. 图书馆数字资源统计标准和应用指南[M]. 北京:国家图书馆出版社,2010:5.

息采集等方式获取馆藏数字资源,但通过图书馆网站上所设置链接而访问的馆外资源,在图书馆没有通过合法的协议(如法定呈缴权)、许可、其他合约或者合作协议来确保图书馆读者使用权的情况下,是不能作为馆藏数字资源的。通过不同方式建设的馆藏数字资源具有不同的特性,表现在资源内容、使用权限等多个方面,本书重点研究的是图书馆通过数字化、网络信息资源采集等方式自主建设的馆藏数字资源的开发与管理。

1.1.2 馆藏数字资源的特点

与传统文献类型相比,数字资源虽然出现时间较晚,但发展非常迅速,在资源形式、来源、存储等方面都与传统资源存在显著的差异,这些差异成为图书馆进行馆藏数字资源开发与管理的重要基础。

(1)数字资源类型丰富,格式多样。馆藏数字资源涵盖电子图书、电子期刊、电子报纸、数据库、音视频资源、网络信息资源等多种类型。从资源来源的角度看,既包括传统文献的数字化,也包括各种类型的原生数字资源,同时还涉及通过特殊形式可利用的其他虚拟馆藏。这些数字资源分布在不同的应用系统中,不但表现形态不同,而且组织方式各异,呈现出多元化资源并存的特征。从资源格式的角度看,数字资源更是丰富多彩,除了传统的关系数据库记录以外,纯文本、Web网页、文档、图像及音频、视频等都是常见的格式,同一种资源的多种格式同时并存的情况更是普遍存在。

(2)与传统文献密切相关。我国图书馆界从20世纪90年代逐步开展数字化建设,近年来不断通过多种途径进行馆藏数字资源开发与建设工作,其中,基于本馆已有传统资源和特色资源而建设的数字资源是一种重要的方式。通过数字化方式建设的这些数字资源在内容上却与图书馆馆藏的传统资源密切相关。

(3)长期保存困难。传统文献由于资源应用方式相对简单,图书馆主要关注与保存相关的物理环境即可,而数字资源的长期保存涉及更多的问题。数字资源一般存储于特殊介质,并且对存取设备有一定

要求,使一些数字资源需要与特定的软件相结合才能读取,实际上关系到一个完整的技术体系,因而增加了保存难度。此外,数字资源长期保存通常与知识产权密切相关,特别是存档权、存储管理权和服务权^①,使长期保存问题的复杂性超出了纯粹的技术领域。

1.1.3 馆藏数字资源的类型

数字资源类型多样,常见的类型有:

(1)从数据的组织形式上看,数字资源包括数据库、电子期刊、电子图书、网页、多媒体资料等。

(2)从存储介质上看,数字资源可分为磁介质和光介质两种类型。其中,磁介质包括软盘、硬盘、磁盘阵列、活动硬盘、U 盘、磁带等类型;光介质包括 CD、DVD、LD 等类型。图书馆常用的数字资源存储介质为硬盘、磁盘阵列、磁带及 CD、DVD、LD 等。

(3)从资源提供者来看,数字资源可分为商业化的数字资源和非商业化的数字资源。前者包括数据库商、出版商和其他机构以商业化方式提供的各种电子资源,如 Elsevier 公司的 SDOS、EBSCO 公司的 Academic Source Premier、中国期刊网等数据库,图书馆需要支付一定的费用后才能提供给一定的读者群,或者读者个人通过读书卡和其他方式购买数据库的使用权。这些数字资源内容丰富、数据量大,是目前图书馆馆藏资源建设中的重要内容。后者主要指机构自建的特色资源库、开放获取资源、机构典藏和其他免费的网络资源,这些资源或者由图书馆自行建设,或者可以从网络上免费获取。当然,图书馆特色资源库在建成之后也可以以商业化方式进行运作,此时,对其他图书馆而言,可以称之为商业化数字资源。

(4)从资源的使用方式上看,数字资源可划分为三大主要类型:一是互联网方式,即需要通过远程连接或网络连接进行访问,主要是指

^① 宛玲,张晓林. 数字资源长期保存过程中的知识产权问题分析[J]. 中国图书馆学报,2005,31(3):65—69.

存储在非本地计算机上的数据库需要通过互联网并使用账号或指定IP段等方式进行远程访问检索等操作,这种方式又被称作网上包库模式;二是局域网方式,即通过本机构内部局域网进行访问检索,它主要指的是数据库系统安装在图书馆的网络服务器上,多用户可同时在图书馆局域网利用数据,这种方式又被称作镜像站点模式;三是光盘检索方式,即购买的数据库以光盘为存储方式,在使用时需通过光盘才能进行检索。随着信息技术的发展,网络实现了信息表达和传输的质的飞跃,光盘检索方式逐渐被互联网方式和镜像站检索方式所取代。

(5)从资源的分布区域上看,图书馆数字资源可划分为两类:一是现实馆藏部分,即在地理位置归属于本地馆藏,是由图书馆自身采集、收藏、整理的数字资源,储存在本馆服务器中,数据所有权归本馆,图书馆具有利用、支配、更新、修改等权利,这一类型的数字资源包括图书馆自建或购买的安装在本地服务器上的各种类型的数据库;二是虚拟馆藏部分,是指通过计算机系统和通信设备在互联网上共享的馆外数字资源,即通过互联网获取只有使用权的数字资源,包括购买或租赁使用的远地数据库和本地服务器实现了连接的网络信息资源。对一个图书馆来说,只要拥有一台与网络相连的计算机,那么网络上所有的信息资源都有可能成为该图书馆馆藏资源。因此,理论上图书馆的虚拟馆藏十分巨大,除本馆现实馆藏以外,整个因特网上的数字资源都可以作为其虚拟馆藏。实际上,现实馆藏和虚拟馆藏的概念是相对的,本馆的现实馆藏,可能同时就是其他馆的虚拟馆藏。

1.1.4 馆藏数字资源开发与管理

随着数字资源成为图书馆资源建设的重要内容,越来越多的图书馆开始以不同形式开展馆藏数字资源开发与管理活动,也取得了丰富的成果。本书在开展相关讨论之前,首先明确界定馆藏数字资源开发与管理的范围,并分析图书馆开展馆藏数字资源开发与管理的优势所在。

1. 馆藏数字资源开发与管理的含义

馆藏数字资源开发是指对纳入馆藏范围的各种类型、文种和载体的馆藏数字资源所含价值进行挖掘,通过多种形式提供给用户,实现资源保护与增值,提升利用价值。馆藏数字资源管理是指对数字资源的采集、加工、发布、服务、存储等生命周期各个环节进行的规划、组织、协调等业务的总称。

(1) 自建数字资源管理

自建数字资源包括馆藏文献数字化、本馆原生数字资源及本馆采集的网络信息资源。自建数字资源可由本馆自主加工完成,也可以外包形式由其他机构完成。自建数字资源建设的各个环节必须严格遵守资源建设标准规范,对尚无相应加工规范的资源,可采取测试数据实验、方法总结、制定规范、批量建设的方法执行。

(2) 引进数字资源管理

引进数字资源是指通过各种方式从本馆以外的权利人(包括团体和个人)处获得使用权或保存权的数字资源,不包括利用本馆网络可以无限制获取的免费网络资源。

数字资源引进方式包括购买、许可授权、受赠、受缴、交换等。对不同类型的数字资源,应选择适宜的引进方式。

数字资源引进工作包括组织试用、实施引进、联系开通、采访信息动态维护、综合评估、统计分析、相关资料的归档和保存等。

(3) 合作建设数字资源管理

合作建设数字资源是指本馆与国内外相关机构联合建设的数字资源。

2. 图书馆馆藏数字资源开发与管理优势

图书馆作为各种类型文献资源的收藏与服务机构,根据数字资源的特点而进行有针对性的开发活动,其优势是显而易见的。

(1) 随着图书馆文献资源的不断累积,图书馆馆藏空间变得越来越拥挤,而数字资源载体的存贮量大、体积小,开发数字形式的馆藏能够节约大量的图书馆空间,尤其有利于保护一些珍本、善本、古籍和历

史文物等文献资料。

(2) 传统文献资源受到物理载体流通的限制,共享范围有限,而数字资源传递距离远、速度快,容易实现大范围共享。数字资源的获取在形式上不受时间、区域的限制,在版权允许的情况下,用户只要有一台计算机与数字图书馆相关的网络连接,就不必亲临图书馆,可在任何时间、地区自由地浏览、查询、检索、存取、利用网上的数字资源,而且不受数量、时间限制,也不受人为因素的影响。

(3) 图书馆拥有海量文献资源,对这些资源的组织与检索是图书馆资源建设工作的重中之重。数字资源可具备强大的组织和检索功能。经过科学合理组织的数字资源能够实现多维度检索,检索点甚至可以包括图像、声音等形式的信息,与传统文献资源相比具有明显优势。数字资源不但便于图书馆对海量文献资料进行组织,同时也有利于用户进行检索和查询,提高图书馆资源的利用效率。

(4) 数字资源独有的特性决定了其具备较好的共建共享条件,图书馆可以灵活运用本馆范围外的各项资源,探索和实践多种建设模式,实现本馆资源建设和服务能力的提升。

1.2 馆藏数字资源开发与管理意义

1.2.1 保护与补充传统馆藏

对于受到损伤的原始文献,其数字拷贝可以起到补偿原件的作用。例如,图书馆馆藏的纸质脆化的文献、脆弱的发黄褪色照片等文献,均可以通过转换成数字文献,从而对原件进行限制性使用。同时,避免或减少用户直接使用原件,将原件存放在适宜保存的环境中,达到既保护原件,又便于利用的目的。对于一些珍贵文献,将其数字化拷贝后提供利用,还可减少或避免原件丢失,并扩大其利用范围。

不但如此,数字化方式还可以改善和提升原始文献的使用条件。例如,在数字化过程中使用一些方法增加原始文献的信息,或通过消

除文献上外来的污迹、斑痕等改善、恢复原始文献。通过这些处理,使原始文献包含的信息更加丰富,显示的效果更好,便于用户查找、检索、研究与处理资料。例如,转换成文本格式的文献,更有利于用户以自然语言进行全文检索与快速存取。

毫无疑问,文献数字化可以产生相对原始文献而言可靠而功能性更强的数字资料,支持新型研究或扩大研究。例如,对象文献数字化后的影像可以支持长达几分钟的定格观察,数字化后的图像资源支持用户利用软件放大功能在屏幕上仔细地鉴别图片,某些图像还可以在高解像力下仔细观察,以利于资源的发现、比较与鉴别。以手稿资源开发为例,原始手稿字迹的辨别一般较为困难,而对手稿扫描并经图像数字放大后,文字研究者便可以在屏幕上更仔细地考察作者笔迹以及研究文献内容。扫描后的手稿,经图像放大后,甚至可以分辨出文字的重叠部分。

1.2.2 提高图书馆管理效率

在网络环境下,建设开发数字资源实际上也是提高图书馆管理效率的手段之一。一是图书馆能够借助自动化工具,方便地进行文献资源建设和提供文献信息服务,不仅可以减少管理原始文献、存取原始文献的成本,而且建设和服务的速度快、效率高;二是将传统文献转化为功能强大的数字资源,不仅可以提升文献存取的功能,而且不再受到文献物理载体流通形式上的限制,有利于提高资源利用率;三是通过数字化馆藏、网络化链接的建设,用户在合理范围内可能克服文献利用时间和空间上的限制,使馆藏文献得到更广泛、更深层的开发利用,更好地满足用户多方位、多功能、多层次的需求;四是有利于实现和深化资源共享,资源共享是当前图书馆文献信息建设与服务的重要模式,数字资源具备实现更大范围、更高效率共享的基本条件。

1.2.3 促进图书馆建设与服务

馆藏数字资源开发与管理是一项艰巨而有意义的任务,随着馆藏

数字资源建设的不断深入,一方面,图书馆的服务对象将大幅增加,服务方式将更加多样化,服务范围将更加广泛,在很大程度上将改变图书馆的服务模式;另一方面,图书馆在资源开发过程中,通过对资源的分析、加工、组织与整合,直接参与数字资源内容建设环节,打破了在传统文献时代图书馆主要参与资源利用环节的格局,直接影响图书馆的战略发展方向,大大推动了图书馆建设模式的转变和创新。不仅如此,在互联网索引系统和搜索引擎技术迅速发展的时代背景下,数字出版如火如荼开展起来,图书馆需要重视馆藏数字资源的开发与管理,依靠自身服务模式和建设模式的转变与创新,在未来知识服务领域中赢得一席之地。

1.3 馆藏数字资源开发与管理的内容

1.3.1 馆藏资源数字化

馆藏资源数字化是图书馆自主开发数据库的基础,也是图书馆数字资源建设的重要内容。目前,全国已有许多图书馆相继开展大规模馆藏资源数字化建设,常见的数字化内容包括古籍资源、民国文献、普通中文图书以及各类馆藏特色资源,图书馆一般将完成加工的数字化全文资料和书目数据上网供读者检索和浏览。

1.3.2 数据库建设

数据库建设是图书馆数字资源建设的核心内容。数据库主要分为自建数据库和购买数据库两部分。图书馆自建数据库主要有书目数据库、特色数据库、专题数据库等。最基本的是书目数据库,将馆藏数字资源纳入 OPAC 目录范围,实现了文献管理和检索的自动化,为读者网上查询创造条件。特色数据库是图书馆数据库建设的重点,也是图书馆充分展示其个性,提高其信息服务竞争力的重要途径,主要类型包括馆藏特色数据库、学科特色数据库、学位论文数据库、地域特

色数据库等。目前,许多图书馆都已建成自己的特色数据库,多数图书馆已为读者提供查询和浏览服务。外购数据库也是图书馆馆藏数字资源建设的重要组成部分,各种商业数据库的主要特点是包含大量的学术资源,利用方便,能够满足读者对专业化信息资源的需求,为读者提供全面、方便和快捷的服务。

1.3.3 网络数字资源开发

因特网信息资源极为丰富,图书馆对它进行开发组织,就可以使这些分布在全球的网络信息资源成为本馆的虚拟馆藏。近年来,利用一些专用软件和技术对原生性网络信息资源进行开发利用已经成为一些图书馆数字资源建设的重要内容。这种开发和组织是根据用户的需求与图书馆资源建设的需要,搜索、选择、挖掘因特网中的信息资源,以一定方式下载到本馆或本地网络之中,经过分类、标引、组织,通过网络或其他方式提供给用户使用,或者链接到图书馆的网页上。由于虚拟馆藏汇集了全球范围内有关专业的信息资源,使图书馆和各类信息机构不仅有丰富的资源保障,而且为扩展其服务功能创造了条件。

网络信息资源开发的主要形式包括学科信息门户、开放获取资源建设、网络信息资源采集等。相对而言,网络信息资源涉及的内容范围比较广泛,也就意味着图书馆具有较大的选择范围,在建设过程中,图书馆需要根据本馆的读者需求与建设条件,选择适当的建设方向以及内容组织方法,以最终形成有别于一般互联网信息的特色化资源。

1.4 馆藏数字资源开发与管理的影响因素

1.4.1 文献资源建设需求

馆藏数字资源开发与管理的要素受到诸多要素的影响,而文献资源建设需求是图书馆进行数字资源开发与管理最为密切的要素。如果目标

读者群不支持使用该文献资源,或者不愿意使用该文献资源的某项功能,那么对其进行数字化形式的开发与建设的必要性评估就要受到影响。另外必须考虑的问题是,图书馆计划进行的数字资源开发项目是否与本馆的馆藏发展与建设政策保持一致,该项数字资源开发工作在当地、国家乃至世界范围的数字资源开发建设中处于何种地位,审慎评估开发项目的必要性和可行性。因此,在选择数字资源开发对象时应参考其他图书馆或国际国内已有的相关数字资源建设规划与项目经验,结合本馆文献资源建设需求制定详细的开发规划方案。

1.4.2 标准化建设

馆藏数字资源开发与管理是一项复杂的系统工程,涉及硬件基础平台、应用软件系统、信息资源建设等多个方面。标准规范作为馆藏数字资源建设的基础,是开发利用与共建共享资源的基本保障,是保证数字资源在整个数字信息环境中可利用、可互操作和可持续发展的基础。因此,在数字资源开发与管理中,应确保开发与建设的标准化。一些主要的数字图书馆建设项目中都建立了一系列的标准规范或指南性文件以指导和规范数字资源建设与服务。如我国开展的“国家数字图书馆工程”、美国国会图书馆开展的“美国记忆”项目等,ISO、NIISO、W3C 等一些国际性标准化组织也致力于数字资源建设标准规范的发展,涉及元数据与对象数据描述、资源组织、资源服务、资源保存等各个方面。关于开放获取资源、新媒体资源等的标准规范也逐渐出台,这些标准规范广泛应用于数字资源建设与服务中,逐渐形成数字图书馆建设的主流标准。尽管数字资源加工与建设的标准规范也在不断地随技术发展而调整,但标准规范的意义在于它必定可以在相当长的时间内保证用户对该数字资源的广泛存取,是影响图书馆馆藏数字资源开发与管理的重要因素。

1.4.3 规范化管理

要使馆藏数字资源开发建设达到预定目标与要求,图书馆必须对