

●刘家真

馆藏文献数字化的原则与方法(上)^{*}

摘 要 馆藏文献数字化具有保护原始文献、提高信息服务效率、利于科学研究等重要作用。拟实施的数字化工程可分为利于提高存取功能的、有助于文献保护的、便于馆藏利用的、可为其他馆藏提供背景的等等类型。不同类型工程的数字化文献类型也不同。馆藏文献数字化应当考虑其是否利于资源建设、市场效益大小、是否符合标准化要求,以及是否利于加强管理,等等。表2。参考文献3。

关键词 馆藏数字化 作用与意义 数字化工程 选择原则

分类号 G250.76

ABSTRACT In the paper, the author thinks that the digitization of library collections plays important roles, such as preserving original documents, improving information services and facilitating researches. The proposed digitization projects can be divided into several types according their purposes, such as for document preservation and for collection utilization. In the digitization, people should consider resources development, market profits, standardization and management. 2 tabs. 3 refs.

KEY WORDS Digitization of library collections. Role and significance. Digitization project. Choice principle.

CLASS NUMBER G250.76

馆藏文献的数字化是信息资源建设的一部分,它可以丰富虚拟图书馆信息资源,促进数字图书馆的发展。馆藏文献数字化涉及到馆藏文献的价值确定,文献的形体特征,当前与今后的用户需求,转换技术的确定、格式选择,怎样描述、传送与归档,以及怎样与其他数字化成果产生联系等许多方面。哈佛大学学者对馆藏文献的数字化决策提出了8个问题。他们认为这些问题中只要一个问题的回答是“no”,就得考虑取消馆藏文献数字化的提议。这些问题是^[1]:

(1)拟数字化文献是否具有足够的实质性价值,以保证数字化工程效益;

(2)该文献数字化后是否可明显地提高存取率;

(3)该数字化工程究竟达到了什么目的;

(4)数字化拷贝满足了确定的需要吗;

(5)数字拷贝的传送符合版权法或得到特许吗;

(6)当前技术产生的高质量图像是否可满足我们的要求;

(7)有足够的经费与充分的技术力量支持吗;

(8)当地的信息基础结构是否充分,可以支持数字文献的传送吗。

可见馆藏文献数字化是一项复杂的工程。不少国家的图书馆、档案馆为了指导本国文献的数字化转换工作,纷纷拿出了自己的指导方案,学者们也针对这一问题展开了研究。笔者在对这类文献与研究报告研读后,提出了自己对馆藏文献数字化原则与方法的思考。

1 馆藏文献数字化的意义

1.1 馆藏文献数字化可以降低原件丢失与损坏的风险

纸质脆化的文献、发黄褪色照片等文献,均可以通过转换成数字文献来对原件进行限制性使用。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,达到既保护了原件,又便于利用的目的。对于一些珍贵文献,利用其数字化拷贝提供利用,可减少或避免原件丢失机会,并扩大其利用的范围。

1.2 受到损伤的原始文献,其数字拷贝可以补偿原件的作用

数字化可以用某种方法增强原始文献信息,或通过消除文献上外来的污迹、斑痕等改善、恢复原始

* 本文为武汉大学重大科研项目研究论文。

文献的信息。馆藏文献数字化还可以解决纸张脆化的文献利用问题,例如,将脆化资料的碎片在扫描仪上拼好,通过扫描仪形成原件的数字化拷贝,使原来无法利用的脆化文件可以被用户存取。

1.3 馆藏文献数字化可以使图书馆更有效率地管理馆藏与提供服务

馆藏文献数字化拷贝可以支持传统图书馆的多种需求,为开展特色服务创造条件;可以较低费用与风险对馆藏文献进行拷贝与馆际互借等。

1.4 馆藏文献数字化可以满足用户对文献的多种需求

馆藏文献数字化后对存取最为有利,它允许用户对某些文献、某些章节进行快速阅读,按需浏览与保持随机存取;它允许用户自己查找、检索、研究与处理资料;它能满足用户共享资源与及时存取资料。例如,转换成文本的文献,便于用户以自然语言进行全文检索与快速存取,等等。

1.5 数字化的馆藏文献可提高原始文献的利用率

传统文献转换成数字格式后,图书馆可以通过在线陈列,充分地为用户介绍馆藏,使用户及时发现自己曾不知道的图书馆资料与有关服务,通过扩大用户对馆藏的知晓程度而促进用户对传统馆藏的利用。所有的数字化产品都易于导航,通过内嵌资料的链接,可增强对原始资料的应用程度与应用范围,使数字化文献比原件更易按需浏览并保持随机存取,因而提高了原始资料的利用率。

1.6 文献数字化可以产生对原件可靠而功能性更强的数字资料,可以支持新型研究或扩大研究对象

文献数字化后的影像可以支持长达几分钟的定格观察,使我们可以利用软件放大功能在屏幕上仔细地鉴别图片,这使得极小的可视资源图像(如照片、图形、图画)能被浏览开发。

某些图像还可以在高解像力下仔细观察,以利于比较、鉴别。例如,手稿字迹鉴别过去极为困难,如今手稿扫描并经图像数字放大后,使文字研究者可以在屏上更仔细地考察作者笔迹。扫描后的手稿,经图像放大后,甚至可以分辨出打字时文字的重叠部分。

1.7 更便于有偿服务

馆藏文献数字化后,可刻录到光盘上自用或商业发行,或是制作网络版放在服务器上供远程检索服务。通过电子商务系统,便于在线有偿服务与订购。

2 馆藏文献数字化工程与拟数字化的文献类型

2.1 利于提高存取功能的数字化工程

该工程的主要目的是期望通过馆藏文献的数字化,提高对原始文献的利用。该工程的拟数字化文献有以下类型:

(1)能满足本馆主要读者需求的文献,或转换为数字拷贝的利用会受到读者广泛支持的文献;

(2)具有重要历史或知识内容的款目,国家珍善本;

(3)相对说来不太为人了解的馆藏文献,通过数字化达到提高存取功效目的;

(4)以某种格式数字化后,可能成为新的馆藏,而使其利用率增加;

(5)数字化后能使原始文献信息价值增殖的馆藏文献;

(6)数字化后有利于克服传统文献的利用障碍,馆藏文献数字化后可解决用户对某些原件存取的障碍。如原件幅面过大或易碎而难以以传统方法存取的文献,等等;

(7)所有的检索工具都可以考虑数字化,特别是当其数字化后可以明显提高对印刷馆藏或缩微胶片馆藏存取的检索工具,以便于提高原始馆藏利用率。

2.2 有助于文献保护的数字化工程

通过馆藏文献数字化,减少用户对原件直接利用而有助于保护馆藏文献的数字化工程,这类工程有以下几类文献可考虑数字化:

(1)处于不稳定载体上的文献;

(2)价值高的文献;

(3)利用率高的文献;

(4)需修复的文献。

2.3 数字化后便于馆藏利用的数字化工程

(1)数字化后容易导航并以数字格式处理的款目;

(2)通过数字化后能增加原始馆藏的使用途径的款目。

2.4 对本馆具有重要意义的数字化工程

(1)数字化后,能满足投资者兴趣,具有吸引投资潜力的文献;

(2)通过数字化拷贝进行有偿服务可以产生效益的文献;

(3)该文献数字化后可以提高本馆声望或知名

度的文献;

(4) 具有吸引或促进其他数字化活动开展
的数字化工程。

2.5 为其他馆藏提供背景的数字化工作

涉及同一主题而分藏在不同图书馆的传统馆藏,由于处于不同地理位置而难以用传统方法对其进行系统研究的馆藏可考虑数字化。分藏不同地区的相关资料,每个馆应分工合作地对自己所收藏的部分文献进行数字化。例如,分工合作地扫描同一主题的照片或分藏不同地区的连续出版物等。

2.6 受资助的馆藏数字化工程

由投资者对自己感兴趣的馆藏进行数字化投资的馆藏数字化工程。

3 馆藏文献数字化应考虑的问题

馆藏文献数字化是一个相当复杂的工程,它涉及到许多环节,例如:选择/准备/转换/编目/存储/处理/传送与长期维护等,这众多环节都是环环相扣,相互影响的。以上环节若不加以全面考虑,或仅仅孤立地强调某一方面因素,例如仅侧重考虑用户需求或过多考虑当前技术能力等而忽略其他相关因素,都可能危及转换后的数字馆藏的长期使用问题。

哈佛大学的 D. Hazen 认为馆藏文献数字化必须至少考虑 20 个方面的问题,它将这 20 个问题排列于表 1^[2]。

这 20 个问题可以概括为以下 4 个方面:

3.1 文献资源建设

馆藏文献数字化之前必须看其是否有利于馆藏建设。如果目标读者群不支持使用该文献资源,就不值得对其进行数字化了。另外必须考虑的问题是,该馆藏的数字化是否与本馆的馆藏建设政策一致,该款目的数字化对当地或国家的数字文献资源建设是否有帮助,甚至应当考虑对国际数字图书馆建设是否有利。这就要求我们在对数字化对象选择时应参考其他馆与国际上已有的回溯资源数字化规划,制定数字化对象的选择策略,使本馆的数字化工程与国家、甚至国际数字图书馆建设相衔接。笔者认为,要做到这一点国家有必要组建数字化指导委员会,对已有的与正在制定的回溯资源数字化工程进行指导与协调,使之利于国家与国际的数字图书馆资源建设,避免重复开发与有目的地开发。

3.2 效益

馆藏文献数字化还要考虑到市场效益。D. Hazen

表 1 馆藏文献数字化必须考虑的问题

	问题 1	问题 2	问题 3	问题 4	问题 5
评估	存在用户支持吗	地方馆藏政策是什么	它有利于国家的利益吗	类似数字化产品在他处已有	目的是什么? 仅为了保护馆藏吗
收益	数字化可减少原件或使用于存取	原文献的知识是否得到了增强	易于导航吗	与其他馆藏链接了吗	是否有利于对已损坏的原件的利用
标准	是否遵循了适当标准	原件有适用的硬件平台支持吗	软件是否易于使用	元数据是否符合已有的标准	归档保存的要求是什么
管理问题	有足够的资金吗	版权与授权问题解决了	本单位有足够的技术力量吗	有商界的合作者吗	所取得的效益与支出是否平衡,合理吗

提出馆藏文献数字化工程必须考虑的第二大因素是是否有效益。他通过 5 个方面的问题来检查数字化工程是否给图书馆带来效益,并认为只有以上 5 个方面均可受益,该数字化工程才是成功的。

3.3 标准化

第三个应当考虑的因素是标准化,尽管标准也在不断地随技术进步而改变,但它必定可以在相当长的时间内保证用户对该数字化产品的广泛存取,还可为今后数字产品的技术迁移带来便利。在标准化问题上,模数转换必须考虑:模数转换所使用的标准是国际标准还是国家标准;产生的数字化资源可否得到各种硬件平台支持;用于递送该数字化产品的软件可否容易得到;元数据是否符合已有的国际标准,如都柏林元数据标准;以及是否考虑到了今后硬件、软件与数据迁移的安全,是否满足了数字信息归档管理的基本条件与要求等。

3.4 管理

要使回溯资源的数字化达到所需目的与要求,必须对有关的管理因素进行认真思考。数字化需要充分的投资,如果有基金资助,还必须使数字化产品满足资助部门的要求。数字化工程还必须对拟数字化文献的版权情况有充分了解,因为版权所有者可以控制该文献的发布权,不支持图书馆免费存取其作品。与版权所有者磋商是要付出费用的,因而在数

表2 模数转换的程序与问题

步 骤	涉及问题
拟数字化对象的选择 ↓	文献价值;版权;原件状况与完整性;数据丢失后,原件可否再扫描;文献被数字化的范围(局部或全部)。
按需组织信息内容 方案制定(如文档 页面图像或可检 索的文本数据库) ↓	适当的文件格式与存储媒体选择。
数字化前的 原件准备 ↓	保留原件吗?(建议保留)原件完整性与强度检查。
品质检查 ↓	需要具有可接受的长期保存质量要求;如果由第3方提供技术服务,是否有文档说明与质量担保。
最终的索引与目标 ↓	资源开发的元数据,管理元数据与数字拷贝的保存。
将数据装填进 计算机 ↓	制作适当的备份拷贝。建议制作4个拷贝,至少有2个是分开保存在不同地理位置。
执行归档与保存 策略,或传送到 保存部门	格式、存储媒体、文档、转换程序均应标准化;母文件与备份的存储;媒体更新以及对付技术淘汰的策略。

字化过程中,还必须考虑存取政策与控制措施、费用

机制与付帐程序等。受版权保护的文献,可使用口令、直接付费、权限控制等措施控制使用;对不必控制使用的回溯文献数字化产品,则可以对阅读、下载与打印以不同价格对不同用户以不同规定加以管理。

在管理问题上还必须考虑,模数转换的费用是否可与取得的效益相平衡,以证明该项馆藏的数字化是合理的。费用/效益评价是十分困难的,但又是极为必要的。据有关资料介绍:十余年来,国家已投入10亿元制作了近千个数据库,但据统计,真正有市场价值的只百余个,能提供商品服务的数据库不到10%。大多数是年年靠财政上的不断输血来维持^[3]。

以上仅仅是从宏观上对馆藏文献数字工程加以考虑,在实际操作中,需要考虑的问题更多更细,表2是数字化处理中必须考虑的有关问题。

由上可见,馆藏文献数字化工程是十分复杂的。数字文献形成后,是否可以达到预期目的,既取决于拟数字化对象的选择,也取决于数字化处理过程的质量控制。

参考文献

1,2 <http://www.thames.rlg.org/preserve/matrix.hemL>
3 董焱,刘兹恒.图书馆馆藏文献数字化.图书情报工作, 2000(7)

刘家真 武汉大学信息管理学院教授,博士生导师。
通讯地址:武汉市。邮编 430072。

(来稿时间 2001-05-08)

(上接第 36 页)

26 刘嘉.元数据之研究.北京大学信息管理系博士论文, 2000:176~179
27 Zhuang Jing.中国试验型数字式图书馆项目. 1 Oct. 1998. Internet WWW page, at URL : < <http://www.nlc.gov.cn/dlib/dcl.htm> > (version current at 15 May 2000)
28 刘炜,赵亮.元数据方案实施意见. 15 June 1998. Internet WWW page, at URL : < <http://www.istis.sh.cn/istis/dlib/report/metadata.html> > (version current at 25 January 2000).
29 中国数字图书馆工程筹备领导小组办公室.“中国数字图书馆工程”筹备领导小组成立,有关研究、论证工作正

式展开.演示文稿,第1期(1998年8月26日). Internet WWW page , at URL : < <http://www.ccnt.gov.cn/digilib/dilib.htm> > (version current at 16 May 2000).
30 清华大学建筑数字图书馆项目组.清华大学建筑数字图书馆. Internet WWW page , at URL : < <http://166.111.68.208/digital/default.asp> > (version current at 7 January 2001).
31 广东省中山图书馆.数字式中文全文文献通用格式:征求意见稿. 2000-01-22

刘嘉 博士,北京大学信息管理系讲师。通讯地址:北京大学。邮编 100871。

(来稿时间:2001-04-16)