

# 古代史料数字化的两种解决方案及其实施

王志轩

(南京大学 历史系, 江苏 南京 210093)

**摘要:**史料可分为历史文献和史迹遗存两类,与之对应的是数字图书馆和数字博物馆两种解决方案以及数字文本和虚拟现实等技术。近年来这两种方案从理论到实践都有了较大发展,而检索技术成为史料数字化的关键所在,史料数字化已成世界潮流。

**关键词:**史料;数字技术;数字图书馆;数字博物馆;虚拟现实

**中图分类号:**K05;TP391 **文献标志码:**A **文章编号:**1002-0292(2006)05-0124-03

史料包括历史文献与史迹遗存,是古代史研究的基础。中国历史悠久,历代文献遗物浩如烟海,近现代又有大批新史料发现与累积。史学界对史料的整理和研究历来非常重视。人类已经进入信息时代,我们现在不仅面临着用新技术来整理旧史料的问题,而且随着考古新发现的不断涌现和科技的发展,史料的容量和形式急剧膨胀。以何种方法处理、存储史料,并且能够让学者从中方便迅速地检索和得到所需的信息,是当前史料学的一个大问题。

数字化是史料处理的必由之路。第一,把各种史料统一为计算机数据,方便存储和检索;第二,使史料的物理体积大大缩小,方便携带;第三,可以低成本无差错拷贝,方便传播和普及。目前,不论是印刷文档、手写稿,还是电子文档、音像文件、甚至于遗址与文物,越来越多的史料逐渐被数字化,出现了数据仓库(Data ware housing)、数据集市(Data mart)、数据采集(Data mining)等新概念和信息数字化处理方法,逐步形成了大数据量存储和管理模式。全球范围内,建立史料数据库逐步成为潮流。联合国推动 Memory of the World 计划以挽救某些濒临灭绝的人类文明,西方各国也无不积极推动数字图书馆、数字博物馆或数字典藏建设,如美国有国会图书馆的 American Memory 计划和六所大学所主导的数字图书馆计划等;加拿大有国家数字典藏计划(Canada's National Digital Collections);日本和 IBM 合作的数字博物馆计划;大英图书馆的敦煌数据数字化计划等。自上世纪 90 年代中期以来,因特网技术对整个社会的影响程度全面加深,信息科技已然启开了社会全面急速变迁的序幕。借助网络这一利器,分散在世界各地的史料数据资源得以有效整合,因特网已经成为一个超大型的史料数据库。但是,从总体来看,史料数字化的研究和实践才刚刚起步,还处于探索阶段。因此,对史料数字化进行整体的研究是很有必要的。本文把史料按照其性质分为两类进行讨论,其一为人类精神产品类史料——文献,其二为物质产品类史料——文物。

## 一、历史文献的数字化

历史文献作为人类精神产品,具有可复制性和可移植性,并且必须依托载体而存在。国家标准局批准公布的 GB4894—85《情报与文献工作词汇——基本术语》定义道:“文献是记录有知识的一切载体”。根据这一标准定义,历史文献就是以文本、图形、符号、音像、软件等各种形式固化在一定物质载体上的历史资料,其物质载体为:纸、胶片、磁带、光盘等。历史文献数字化的目标是将原来在各种物理载体上

收稿日期:2006-03-06

作者简介:王志轩(1976-),男,南京大学历史系博士生。

存储的信息,统一转变为最基本的比特,用0和1来表示,成为计算机可以识别的数据。

历史文献数字化的终极形式为数字图书馆的出现。所谓数字图书馆,就是利用现代信息技术对有价值的图像、文本、语音、音响、影像、影视、软件和科学数据库等多媒体信息进行收集,组织规范性的加工和压缩处理,使其转化为数字信息,然后通过计算机技术进行高质量保存和管理,并通过网络通讯技术进行高效、经济地传播、接收,使用户可以在任何时间、任何地点,都能从网上得到借阅服务,成为国家的知识基础设施。<sup>[1]</sup>

建立数字图书馆的关键在于以统一的标准和规范为基础,表现为同一类型文件格式的统一。当前,数字图片格式主要有BMP, GIF, JPG, PCX, TGA, TIF等;数字音频格式有WAV, MP3, MIDI等;数字视频格式有mpg, mov, avi, rm等;网页格式有html, vml, xml, sgml等;数据库有Oracle, SQL Server, Sybase, DB2, Informix, Access等。文献数字化的过程中一般应采取最通用的文件格式,这样既便于为最多的人服务,又便于数据的升级。数字图书实现形式主要有两种:一是采用扫描录入方式将书籍资料按原貌逐页存储为图像文件;二是以文本方式存储文献内容,辅之以全文检索系统构成全文检索数据库。一般来讲,前一种方式适合古籍善本和以往出版图书的数字化,后一种适合新出版图书的数字化。后一种方式是未来数字化图书馆的发展方向。例如:中国社科文献出版社2005年6月宣布,已经建成《说文解字》文本数据库,该数据库采用四字节处理技术,可解决七万多汉字的录入和查询问题。这是中国古籍数字化的一个里程碑。

建立数字图书馆的设想首先由美国人于1994年提出。同年10月,美国国会图书馆推出数字化项目,领导与协调全国的公共图书馆、研究图书馆,将其收藏的图书、绘画、手稿、照片等转换成高清晰度的数字化图像并存储起来,通过互联网供公众利用。继美国之后,英国、法国、日本、德国等西方发达国家以及亚洲的新加坡、韩国也先后提出各自的数字图书馆计划,投入巨额实施资金。我国数字图书馆建设开始于1997年,北京图书馆作为国家图书馆,经国家计委批准建立了“中国试验型数字图书馆”项目,并开展了与此有关的专题科研。同年12月,我国民营计算机高科技企业北京世纪超星信息技术有限公司将其远程图书浏览器安装到瑞得在线网站上,创建了国内首家以图像存储为主的数字图书馆——超星数字图书馆,此后,国家连续实施了“中国数字图书馆工程”和“中国数字图书馆示范工程”,国内各大图书馆和高校纷纷跟进,掀起了一股数字化潮流。据不完全统计,目前国内的中文数字图书馆已达数百家之多,以学术文献为主要内容的超星数字图书馆的数字图书已超过40万种,中国数字图书馆也已达到20万种以上。此外,书生之家“中国图书网”和方正apabi数字图书馆也都非常有实力。

目前,我国科技期刊的全文数据库主要有“中国期刊网”、“万方数字化期刊网”和“维普中文科技期刊全文数据库”等,以中国期刊网为例,收有几年来5300种期刊的全部文章,重要期刊几乎被网罗殆尽。之外,因特网上一些小型的数字文献书库更是数不胜数,如香港中文大学中国文化研究所“华夏文库”及“古文献资料库”;台湾的汉学研究中心典藏目录及数据库,包括典藏国际汉学博士论文摘要资料库、明人文集联合目录及篇目索引资料库、两汉诸子研究论著目录资料库、经学研究论著目录资料库、敦煌学研究论著目录资料库、外文期刊汉学论著目次资料库、国际汉学资源机构录资料库,可进行资料库整合查询;北京书同文数字化技术有限公司的《四库全书》、《四部丛刊》、《康熙字典》的电子版等等。<sup>[2]</sup>千万不要小看这些局部的文献数字化工作,集腋成裘,聚沙成塔,这些数据经因特网而整合为一体,完全不受其物理距离的影响,从长远看,分散的学术单位与学者是文献数字化的真正主力军,整个因特网,就是一个大的数字图书馆集合。

## 二、文物史料的数字化

文物为人类有形的史迹遗存,它具有不可复制性和唯一性。文物作为史料的价值,在于其中包含的历史信息。所以对文物信息的撷取与整理一直受到人们的重视。如对甲骨文的墨拓、整理而成的《甲骨文合集》,对青铜器进行整理而成的《两周金文辞大系图录考释》,这些信息撷取成果目前大多以文献的形式而存在。现代以来,经科学考古而发现的遗址和遗物都有详细的考古报告。报告中的绘图形式有平面图、剖面图、结构图、局部图、分解图、透视图以及器物绘画等;照相有地貌全景、发掘区全景、不同角度和感光的遗物特写,其他如专业技术鉴定报告、文物修复与模型制作、拓印图片等等,有的还附有航拍

图片和电影、录像。考古报告是对文物环境、外观、结构、尺寸、颜色、纹理、成分、气味等信息的全面摄取。文物的数字化即为文物信息的数字化。

文物数字化的最终模式为数字化博物馆,最新技术为虚拟现实(virtual reality)。虚拟现实技术是上世纪90年代以来为科学界和工程界所广泛关注的技术。它利用计算机生成一种模拟的实物与环境,通过多种传感设备使用户“投入”到该环境中,实现用户与该环境直接自然交互。文物的虚拟现实只是静态实物的虚拟,在技术实现上要容易得多,我们把所有依据科学考古而得到的文物信息输入计算机,就能借助虚拟现实生成程序来“制造”出三维的虚拟现实文物,借助这一技术,许多不易接触、难得一见的历史文物可以被普通学者摩挲于股掌之间。另外,使用360度全景摄影技术拍摄的高质量图像也可以生成逼真的虚拟情景,从而使得用户可以使用普通微机通过鼠标和键盘真实地感受到和虚拟现实一样效果,不过它的人机交互性能要差许多。在日本早稻田大学网站,我们可以看到该校建筑学系编纂的从1996年到2002年对埃及Pyramids of Dahshur的8次考古发掘报告,其中包括了用vml技术制作的虚拟现实墓室结构和用QuickTimeVR Panorama全景摄影技术制作的金字塔外部环境。

我国的数字博物馆建设和文物数字化工作已经开始起步,故宫博物院是具有600年历史、8000座殿宇楼阁、百万件文物的我国最大博物馆,其网站于2001年7月16日开通,最初只收录了各类图片4000余张,实现了文物影像的细节浏览和虚拟展厅的360度旋转展示<sup>[3]</sup>。到了2002年7月30日,由中国故宫博物院和日本凸版印刷公司共同建立的故宫文化遗产数字化应用研究所,宣布以故宫太和殿为中心的紫禁城虚拟现实工程目前进展顺利<sup>[4]</sup>。这说明故宫文物的数字化工作已经进入了虚拟现实的实质性阶段。北京四方台数字技术有限公司使用360度全景摄影技术,已经制作了所有中国世界遗产的虚拟实景,此外他们还制作了中国古镇民居、中国历史文化名城、寺庙及宗教建筑等虚拟实景专题。台湾近年启动了“数位典藏国家型科技计划”,其专题内容包括:“台湾建筑史”、“不朽的殿堂——汉代的墓葬与文化”、“玄奘西域行”、“商王大墓重现”、“火器与明清战争”等等,在实施过程中,虚拟现实就是其中的一种重要技术手段<sup>[5]</sup>。使用虚拟现实技术还可以重现已经不存在的历史人物、建筑、地理环境等。2002年1月,中国刑警学院法医系教授赵成文,依据古长沙国丞相夫人辛追颅骨的X光片、出土时拍摄的面部照片,以及长沙市马王堆汉墓帛画和相关历史文献资料,使用自己研制的警星CCK-3型人像模拟系统成功地复原了这一2000多年前的人物相貌,虽然这一技术还有待完善,但是不能否认它很好的发展前景。总体看来,数字博物馆的建设正在受到人们的重视,而虚拟现实技术也处于发展与普及过程之中。

在信息技术高速发展的今天,史科学正在孕育并发生着重大的变革。正在发展形成中的新史科学有四个基本特点已经初露端倪:一、文献多媒体化;二、文物虚拟现实化;三、史料可以廉价复制,极大丰富;四、计算机网络检索技术成为史科学的命脉。史料数字化的时代已经到来,我们正站在新的起点上。

#### 参考文献:

- [1]2002年数字图书馆国际论坛及工程项目洽谈会网站.数字图书馆的历史与演变[EB/OL].<http://www.dlibforum.com/information1.htm>
- [2]吴宣德.网路数字化文献资源[EB/OL].<http://myqf.diy.163.com/>
- [3]苏敏.故宫走进数字时代[N].中国青年报,2001-07-17.
- [4]张焕利.文化遗产数字化应用 故宫虚拟工程进展顺利[DB/OL].<http://202.84.17.73:7777/Detail.wct?RecID=1&SelectID=1&ChannelID=4255&Page=1>
- [5]数位典藏国家型科技计划简介[EB/OL].<http://www.ndap.org.tw/>

(责任编辑 薛正昌)