

# 西部地区地方文献资源数字化建设研究

□ 郭俊平 张桂荣 / 内蒙古工业大学图书馆 呼和浩特 010051

**摘要:** 文章在广泛调研的基础上,通过对西部地区地方文献资源数字化的作用进行分析,阐述了西部地区地方文献资源数字化建设的必要性和可行性,提出了地方文献资源数字化的建设方案,探讨了西部少数民族地方文献资源数字化的方法和途径。

**关键词:** 地方文献, 数字化建设, DC, 元数据, 西部大开发

DOI: 10.3772/j.issn.1673—2286.2011.11.009

## 1 引言

西部地区地方文献资源是指记载西部地区的政治、经济、文化、艺术、人物、历史、山川、气候、风俗、物产、自然资源等方面各种信息的载体。它所涉及的内容范围广,时间跨度大,是关于西部地区最珍贵、最全面的文献,在整个西部地区文献信息资源中占有特殊的地位,其具有的鲜明的地域特色,对西部地区的政治、经济、文化教育等的发展有着极为重要的价值。

西部地区地方文献资源数字化,不仅达到了保护地方文献的目的,更有利于地方文献的传播与开发利用,使其在数字化建设中具有较强的资源优势和竞争力,在西部大开发中发挥更大作用。

## 2 西部地区地方文献资源数字化建设的必要性和可行性

### 2.1 西部地区地方文献资源数字化建设的必要性

西部地区地方文献资源数字化顺应了当今数字技术和文献信息工作的发展趋势,加强西部地方文献资源的数字化建设是非常必要的。

#### 2.1.1 西部地区地方文献资源数字化是西部大开发深入发展的客观需求

西部地方文献记录了西部地区和西部各民族的政治、经济、文化、历史等人文与自然状况,由于其地域性与民族性,这类文献具有很强的使用价值和保存价值。西部地方文献只有加速数字化进程,才能适应西部大开发不断增长的信息需求,在竞争激烈的信息社会中发挥出西部地方文献的特色和优势。

#### 2.1.2 西部地区地方文献资源数字化是保护地方文化遗产的有效措施

地方文献资源是一笔非常宝贵的精神和物质财富,因此,在充分挖掘地方文献资源的文化价值和史料价值的基础上,还必须考虑地方文献资源的长久保存问题。数字化信息十分有利于贮存、传输、查阅和复制。通过数字化,可以减少对地方文献原始载体的直接利用,避免珍贵地方文献资料的被损及遗失现象,有利于它的长久保存,并进一步保护和实现它的文化价值。

#### 2.1.3 西部地区地方文献资源数字化是提高原始文献利用率的重要手段

地方文献资源地数字化能极大地提高文献资源的利用率,一方面可以在保护原始文献的基础上,通过数字化处理,使原始文献的信息价值增值,产生功能更强的数字化资料;另一方面,数字化的地方文献信息可以拓展新的服务领域,可刻录成光盘或制作成网

络版供远程检索服务。正是由于有了这些数字化的地方文献信息资源,地方文献资源的高速检索才得以最终实现,从而促进西部地区地方文献资源整体价值的提升、利用率的提高。

### 2.1.4 西部地区地方文献资源数字化是文献资源共享的发展趋势

用数字技术手段,实现地方文献数字化是资源共享的必然要求,建立地方文献数据库、全文检索系统,实现联机检索,以及网络上的馆际互借、传递、交换等,在形式和手段上为开发利用西部地方文献资源提供了良好的操作平台,为地方文献资源共享奠定坚实的基础,同时,也极大地加快了开发利用西部地方文献的步伐。

## 2.2 西部地区地方文献资源数字化建设的可行性

### 2.2.1 得天独厚的文献资源优势为地方文献资源的数字化奠定了坚实的物质基础

西部地区历史悠久,民族众多,这里不仅有独特的蜀文化、藏文化、滇文化、蒙古文化及夜郎文化等,还有遍布各地的丰富历史古迹,例如敦煌、三星堆、西夏、吐蕃、大理、西凉等众多的古国遗存,还有历代的珍稀文献和其他各种类型文献。同时,西部又是多民族地区,民族文化历史悠久,每个民族的历史、迁徙、民俗风情都以特定的形式保存和流传下来。这些文化、古迹、民俗风情及文献资源,都是巨大的物质财富,它一旦被合理地开发和利用,变成数字化产品,将会产生不可估量的经济价值和社会效应,为西部地方文献数字化的实施打下良好的物质基础。

### 2.2.2 先进的计算机技术、信息存储技术和通讯技术为地方文献资源的数字化提供了必要的技术手段

地方文献收藏机构自动化建设的迅速发展,以及计算机技术、信息存储技术和通讯技术的广泛应用、民族文字信息技术的成功开发研制,为西部地区地方

文献资源数字化提供了必要的技术手段和技术环境。

### 2.2.3 西部大开发为地方文献资源的数字化建设营造了良好的发展环境

西部大开发不仅促进了西部地区经济、社会的发展,而且对地方文献资源的挖掘整理和开发利用工作提出了更高的要求。西部大开发需要地方文献资源的有力支撑,西部地区对地方文献数字化资源需求的不断提高,是西部地区地方文献资源数字化发展的根本动力,同时也为地方文献资源数字化提供了千载难逢的发展机遇和营造了良好的发展环境。

## 3 西部地区地方文献资源数字化建设方案

地方文献资源数字化的建设应在标准化数据的支持下高起点、有步骤、分阶段实施。要先易后难,先建地方文献书目数据库,再建综合型、专题性数据库,技术力量成熟后,建设全文检索数据库和多媒体数据库。

### 3.1 建立西部地区地方文献资源书目数据库

西部地区的文献信息机构在进行地方文献资源数字化、建立各种索引及书目数据库时,要对西部地区地方文献资源进行全面科学的评价,统筹规划、优势互补,以避免重复建设和资源浪费。通过建立各种索引及书目数据库,为Internet用户提供操作简单、功能强大的数字化文献检索服务,为中外学者研究利用西部地区地方文献信息资源提供有效的服务,为广域网下联合目录数据库的建立奠定基础。

### 3.2 建立西部地区地方文献资源联合目录数据库

通过联合建库,实现文献资源共享,是数据库建设的重要途径之一。西部地区地处祖国边疆,又是少数民族聚居区,多年来各单位经费投入不足,计算机应用基础不同,发展水平不平衡,应依据具体情况和本地区社会发展的实际需求,走联合建库的道路,建

立独具地方特色的西部地区地方文献资源数据库保障体系,发挥整体功能,实现信息资源的快速、高效共享。

在西部地区地方文献资源开发中,共用一个软件系统来建设数据库,使数据库标准、规范,数据具有通用性,有利于实现地方文献资源共享。例如,“西北地方文献资源数据库”是以西北五省(区)各类型图书馆馆藏西北地方文献为基础,以西部大开发政策为导向,依托文献信息数字化技术建立的,其目的是将静态的地方文献转化为动态的信息,为西部大开发提供网上信息保障,为西部地区经济发展和社会进步提供智力支持。

### 3.3 建立西部地区地方文献资源专题全文数据库

在书目数据库建设的基础上开发全文数据库,是网络环境下地方文献资源数字化的必然趋势。在西部地区地方文献资源数字化建设过程中,应有计划、有步骤地把西部地区地方文献资源中那些利用率高、学术价值和参考价值大的地方文献资源转化为数字化的可检索的全文数据库。

### 3.4 建立西部地区地方文献资源多媒体数据库

借助计算机等技术手段,通过对音频、视频、图像、文字等资料进行数字化转换、压缩、编辑等技术处理,可以把西部地区地方文献资源中各种声像资料转换成计算机可识别的数字化信息,建成集文字、图片、音像于一体的西部地方文献多媒体数据库,储存在网络服务器上,进行传播,既方便了读者的使用,也有利于对非书资料型的西部地区地方文献资源的保护。

### 3.5 建立西部地区少数民族地方文献资源数据库

西部少数民族地区收藏着丰富的民族地方文献资源,它们集民族性、地区性和资料性于一体,为西部大开发提供了重要的信息资源保障;通过建设少数民族地方文献资源数据库,实现民族地方文献资源的数字化,形成可利用的有序化、规范化的数字资源。

例如,由内蒙古大学图书馆开发的蒙古学特色库已被CALIS中心列为重点建设的民族地方文献数据库。

## 4 西部地区地方文献资源数字化建设的相关标准与制作方法

### 4.1 地方文献资源数字化的软件系统

在西部地区地方文献资源的数字化建设中,选择质量好、性能高的数字化系统软件是首先遇到的一个关键问题。软件的选择应有一定的标准和要求,主要考虑以下因素:要选择那些技术成熟、兼容性强,能与国家数字图书馆工程相连接的软件系统;系统之间标准的统一直接影响到资源库的制作质量和查询服务效果,要选择标准化程度高的软件系统;软件系统应操作简单,界面友好,能将文献书目、全文、图像、声音、动画等输入数据库,并提供一站式的集文字、声音、图像、动画于一体的多媒体检索平台;整个系统应包括四个部分:网页部分、书目数据库部分、数字化档案部分、分散式检索系统(Z39.50)。

国内文献数字技术飞速发展,市场上已有多家文献数字化处理系统出现,如DIPS数字文献处理系统、TRS系统、Adobe Acrobat Reader系统以及北大方正的Apabi系统等。内蒙古大学图书馆的蒙古学特色文献数字化选用了Apabi系统,该系统除了数据加工系统、全文OCR软件、安全发布及管理系统、全文搜索引擎、元数据转换软件、阅读软件等功能模块外,还有适合蒙古文文献数字化的蒙古文排版及网络传播模块(包括蒙古文书版、蒙古文字库、蒙古文输入法等),可提供蒙古文排版及输入,符合UNICODE标准,可以实现蒙古文文献的数字化及网络传播。

### 4.2 地方文献资源数字化的相关标准规范

地方文献资源数字化过程中会涉及很多相关标准和规范,这些问题解决的好坏,直接关系到地方文献资源数字化的成功与否。就目前研究状况来看,地方文献资源数字化的标准规范有数字内容创建标准、数字对象描述标准、数字资源的组织检索标准、数字资源的管理和服务以及长久保存等标准规范。目前投入试验的主要有内容编码方面采用通用的8位编码字符集(UNICODE, ISO/IEC 10646, Universal Multiple-Octet

Coded Character Set)。在标准通用标记语言方面,主要有SGML及其相关标准,HTML是专门用来为WWW建立超媒体文件的语言,扩展的标记语言XML是用来定义处理元数据的基本原则。元数据作为描述数字对象的数据,是所有数字信息资源建设项目的重要基础,需要规定描述数字对象的原则和基本方法,或者在具体范围内规定实际应用的元数据标准与规范。Z39.50是应用服务和检索协议,它是用于网络上不同系统之间互联、互访的协议。地方文献资源中的少数民族文字的编码问题是数字化建设的难点,内蒙古大学图书馆与有关单位合作,经过多年的不懈努力,基本解决了基于UNICODE的蒙古文字符集编码问题,使蒙古文文献的数字化成为现实。

### 4.3 DC元数据在地方文献资源数字化中的应用

DC元数据全称都柏林核心元素集(Dublin Core Element Set),是1995年国际组织Dublin Core Metadata Initiative拟定的用于表示电子信息资源的一种简要目录模式。目前,它由15个数据单元组成。

地方文献资源不但包括传统印刷型资料,还包括非书资料,如金石拓片、缩微品、照片底片、音像资料、网上的地方信息、内部出版物等。由于DC支持对任何内容的资源进行描述,它能比较全面地概括文献的主要特征,所以比较符合地方文献资源的描述要求。DC用于地方文献资源的著录,具有适用范围广、揭示内容全、标引深度大、检索效率高、著录简洁、易于掌握等特点。因此,DC元数据适用于地方文献的组织与揭示,能够满足各类地方文献资源数字化的需要。

与传统的MARC相比较,用DC元数据描述数字化的地方文献资源的主要优势有:

(1) DC元数据是专门为描述数字化资源、网络资源设计的,其15个基本元素全面地揭示了数字化资源的主要特征,涵盖了资源的重要检索点、辅助检索点或关联检索点,以及各种有价值的说明类信息。

(2) DC元数据体系简单,容易掌握。MARC格式体系严谨、结构复杂,一般要受过专门培训的编目人员才能使用,而DC的使用比较容易,非专业编目人员稍加培训即能用其完成电子资源的编目,可以大大降低数字化地方文献资源的编目负荷,节省人力物力,

使其能更多地投入到地方文献资源内容与实体资源的处理以及资源深层价值的挖掘工作中,有效降低地方文献资源的数字化成本。

(3) 西部地区地方文献资源中包含了大量政治、经济、历史、地理、自然资源、科学技术和文教卫生等各个领域的重要信息,这些信息的检索入口词往往是带有浓郁地方民族特色的词汇,有些作为自由词、关键词而未被收入到《中国分类主题词表》。MARC规定使用主题词表里的主题词进行标引,地方文献的这些自由词等只能增设其他相关字段标引。而DC除了使用主题词外还允许使用自由词、关键词进行标引,而且不对标引深度作限制。这些不但对充分揭示地方文献资源的内容特征、满足用户检索习惯、提高查准率有重要作用,同时也符合文献检索从受控词检索向自然语言检索转化的趋势。

(4) Web服务已成为数字化文献资源发布、应用的主要平台。DC记录(HTML、XML)直接就是标准的网络化产品,因此能很好地支持Web服务,也易于使用系统软件或通过开发工具提供的接口进行系统开发。而MARC仅是异构编目系统进行外部数据交换的格式,既不是系统内部数据结构的格式,也不是标准的网络化产品。MARC数据对Web应用的支持需要第三方软件来实现MARC格式到其他Web支持数据格式的转换。

通过以上对比分析,我们不难看出,对于多种类型的西部地区地方文献资源的描述,MARC目前还只适合于书目数据库的著录,而对于地方文献中的实物、网页上的地方信息、全文信息等,MARC格式还带有一定的局限性。而DC元数据则支持对任何内容的资源进行描述。它能比较全面地概括地方文献资源的主要特征,符合《中国文献编目规则》的规范要求,对文献内容的标引深度没有限制,这种面向内容的描述符合地方文献资源的要求,在地方文献资源著录方面值得推广。随着网络中地方文献信息资源的不断增加,地方文献资源专题、全文等数据库的不断建设,DC元数据的描述方式将在地方文献资源数字化建设中得到更加广泛的应用。

### 4.4 地方文献资源数字化的基本方法

地方文献资源数字化的基本方法归纳起来有以下几种:

#### 4.4.1 文字档制作方法

将纸质印刷品的地方文献资源，通过全文录入或计算机扫描以数码形式保存到计算机内。如果是由电脑打字形成的文字档（如地方志），其本身就是一种数字化的文字档。如果文字资料经过扫描后，会先存储成图片档，通过OCR识别，将图片变成文字即成为文字档，然后在原文字的基础上进行编辑、排版、标引等深加工，使大量地方文献资源制作为数字化资源。

#### 4.4.2 图片图形档制作方法

西部地区的地方文献资源形式多样、语种复杂，有些文献可采取图片图形档制作方法。如手稿、碑帖、族谱、家谱、画卷等资料，可采用拍照、摄像等方式转换成图像文件。如果是照片、图片、地图等资料，则只能扫描成图片型或以转换的方式变换为数码化的图形存贮格式，以BMP、JPG、GIF等多种形式保存在光盘或海量存储器上，通过网络实现资源共享。将不同文种的地方古籍文献数字化时可同时处理为文本文件和图形文件两种形式，读者不但可以阅读文献内容，还可以通过图形文件考证地方文献的版本、笔迹等。以图片存储的方式作原版显示，并编入必要的检索目录，这种方法虽然占用较大的存储空间，却能反映民族地方文献的本来面目。

#### 4.4.3 声像数字档制作方法

音像资料可以通过专用设备和相应软件转换成计算机中的声音和图像文件保存在光盘上或网络服务器的存储器内。地方民谣、名人演讲、有保存价值的录音带或光盘等声音资料需要使用声音处理软件，制作成声音型数字档。如果是录像带等资料，则需要以视听软件制作成影像型数字档。

### 4.5 少数民族文字文献资源数字化方法

少数民族文字文献资源数字化通常有两种方式，一是扫描版数字化产品制作方法；二是全息版数字化产品制作办法，这也是所有文献数字化的一般方法。

#### 4.5.1 扫描版数字化方法

其制作流程包括如下步骤：

- (1) 扫描。这是文献数字化的基础。
- (2) 检查。即对扫描的图像进行校验检查。如发现与原文献失真之处，或者漏页、错页，需要进行重新扫描或修复。
- (3) 制作。通过对扫描的图像适当加工制作成所需的数字化文献。通常都使用特定的应用程序或者软件。
- (4) 标引。对做完的数字化文献进行标引，通常采用元数据格式（可以自定义），如果该少数民族文字信息处理标引语言已经比较成熟，可以使用民族语言标引，否则可以使用其他语言（如汉语、英语等）标引。
- (5) 总校。是对整个制作过程的全面检查，检查有问题的需重新修改，通过的则可以发布。

#### 4.5.2 全息版数字化方法

如果能制作全息版数字产品，当然最好。首先，全息版数字化产品内容以文字字符为主，不像扫描版图像文件，占用空间大；其次，全息版数字化产品中的文字全文可以编辑检索，方便用户检索与利用。

其制作流程包括如下步骤：

- (1) 选择制作源。全息版数字化产品的制作源可以是出版社的电子书稿，也可以通过OCR对扫描文献的图像识别而来。
- (2) 制作数字文献。同扫描版。
- (3) 标引。相对扫描版来说容易得多，标引内容可以从文献内部获得。
- (4) 总校。同扫描版。

#### 4.5.3 扫描版与全息版数字化方法优缺点对比（见表1）

### 5 西部地区地方文献资源数字化建设应注意的问题

西部地区地方文献资源的数字化具有一定的复杂性和艰巨性，在建设时要注意解决好以下问题。

表1 扫描版与全息版数字化方法优缺点对比

	扫描版数字化	全息版数字化
开发难度	*	*****
产品体积	*****	*
易用度	**	*****
投资量	**	*****

说明：此处所用比较单位皆为程度，以\*表示，一个\*为最小，五个\*为最大。

### 5.1 西部地区地方文献资源的数字化必须符合西部大开发的需求

随着西部大开发战略实施向纵深发展，信息基础设施建设已经引起社会各界尤其是政府决策层的高度重视。西部地区地方文献资源的数字化建设要特别注意与西部传统物能资源和非传统物能资源相匹配，通过数字化对各种资源进行优化配置，大力提高西部地区地方文献资源的信息附加值。在西部大开发中，西部地区的历史、人文资源和旅游资源有着非常重要的作用，亟待开发和转化成数字化资源，并进一步转化为新型的物能资源。因此，要根据西部大开发的实际需要，努力实现西部地区传统地方文献资源与数字化信息资源的优化配置，从而带动整个西部地区社会政治、经济、文化的发展。

### 5.2 西部地区地方文献资源数字化系统的标准及建设原则问题

规范、标准可以说是数据库的生命，也是数字化建设的起点和基础。西部地区地方文献资源的数字化应该参照国际上通用的置标语言、元数据标准、对象数据标准，结合本地区地方文献资源的特点，选择标准化高、兼容性好、能与各种数字化工程相连接的

软件系统。我国目前正在开发数字式中文全文通用格式，它可以对普通图书、科技报告、学术论文、古籍拓片、地图资料、乐谱、录音资料、影像资料等进行全文著录。关于少数民族语言文字文献数字化标准和原则的制定也要符合国际、国内标准，作为民族文字文献数字化共建共享体系的成员应该密切关注该领域的发展和最新研究成果，尽量应用最先进的标准，避免走弯路。

### 5.3 西部地区地方文献资源数字化建设中的共建共享和统筹规划问题

首先，要避免地方文献资源的重复开发。另外，要保证数字化成果的共享。世界各国在进行数字化建设时，都进行联盟建设，联盟在数字化领域合作的主要目的就是为共享数字化建设的成果，以达到区域性网络资源有效配置的目的。其次，西部地区地方文献资源的数字化建设必须突出地域特色和民族特色，没有特色就没有优势，就很难被市场认同。

## 6 结语

应该说，西部地区地方文献资源数字化工作还处于初级阶段，为了使地方文献资源的数字化建设真正达到预期目的，使投入与获得的价值一致，还必须做好组织决策、技术研制开发等工作，要不断探索前进，加速西部地区地方文献资源的数字化、信息化、网络化建设步伐。根据市场需求，特别要重视西部地区地方文献资源数字化增值产品的开发，在有限条件下，充分利用群体作战和市场合作等优势，利用网络的强大功能，以尽可能小的投入发挥最大的整体效益，共创西部地区地方文献资源数字化的品牌形象，支持西部大开发顺利进行。

## 参考文献

- [1] 姚宏伟. 地方文献数字化及在我馆实践的探讨[J]. 图书馆学研究, 2004(1):13-15.
- [2] 魏文晖. DC元数据在特种地方文献著录中的应用[C]//2002年两南会议论文:126-128.
- [3] 刘泳洁, 王月娥. 关于西部民族地方文献数字化的几点思考[J]. 大学图书馆学报, 2002(4):77-78.
- [4] 雷云. 试论西部地区图书馆特色文献资源的数字化建设[J]. 图书馆学刊, 2002(2):47-49.
- [5] 包和平, 宛文红. 西部图书馆特色数据库现状调查分析[J]. 图书馆论坛, 2003(10):53-54.
- [6] 沙其敏. 地方历史文献存取、检索的趋势以及遇到的问题[J]. 国家图书馆学刊, 2005(1):12-16.
- [7] 肖玲. 谈地方文献数字化资源建设[J]. 图书馆学刊, 2006(2):65-66.
- [8] 龚菲. 我馆民族地方文献资源建设工作研究[J]. 图书馆, 2006(1):100-102.
- [9] 张桂荣. 民族地区建设数字图书馆的方法和途径[J]. 图书馆工作与研究, 2007(2):61-62.

## 作者简介

郭俊平 (1966-), 男, 1988年毕业于北京师范大学图书馆学系, 现任内蒙古工业大学图书馆馆长、内蒙古自治区图书馆学会常务理事兼学术研究委员会主任、内蒙古自治区高等学校图书馆情报工作委员会副秘书长。E-mail: libgipin@imut.edu.cn

## Research of Document Resources Digitalization in Western Areas

Guo Junping, Zhang Guirong / Inner Mongolia University of Technology Library, Hohhot, 010051

Abstract: The paper analyses the effects of document resource digitization in the western areas based on extensive survey, and states document resource digitization development's necessity and feasibility in the western areas, and researches into methods and approaches of document resource digitization of the western minority nationality parts.

Keywords: Local literature, Digitalized construction, DC, Metadata, Great exploit in the western areas

(收稿日期: 2011-06-15)