

严 彬

电子出版物 CNMARC 著录数据的实际应用

摘 要 在进行电子出版物的 CMMARC 字段著录时,对应选择的字段、著录内容及出现的特殊情况等作科学处理,以利于图书馆对电子出版物的管理和检索。参考文献 5。

关键词 电子出版物 中文机读目录 编目

分类号 G254.36

ABSTRACT In the paper, the author thinks that when describing electronic publications in CNMARC fields, people should scientifically treat special situations of fields and contents, so that they can be easily used in the management and retrieval of electronic publications. 5 refs.

KEY WORDS Electronic publications. CNMARC. Cataloging.

CLASS NUMBER G254.36

1 概述

在以藏书整理为核心,以流通利用为目的的现今大多数图书馆建馆方向上,电子版文献的出现仍然包含两种含义:一是对电子出版物本身的收藏与整理,它作为一种新型的文献载体如同石、木、骨、帛、纸等载体一样记录文字信息,具有收藏和研究价值;二是流通与利用,它比以往那些信息载体更具携带方便、检索快捷、存储量大等优点,因而更能发挥其价值。图书馆应适应社会发展需求,发挥电子版文献的作用,围绕对文献的收藏与整理这个核心,加快对电子出版物编目的研究与利用的基础工作。

对一个图书馆、一个读者来讲,数字信息不可能长期占用计算机上的硬盘资源,最终的收藏品还是电子出版物,无论其是以光盘形式还是以硬盘形式,甚至以后出现更大容量的存储产品。对这样的电子版文献如何进行搜集、存放和检索,要有一套编目的标准,这就提出一个如何编目的问题。我们把这样一种电子版文献称作电子出版物^[1,2],并且把其覆盖的范围进一步扩大,从印刷版向电子版的转换,原电子版文献的复制,数据源网站上的文献下载,以及数据源文献的存储和发布等等都视为一种出版的方式,再把电子版上的数据称之为文献,是因其信息是有序的逻辑符号,可以通过一定的人为加工,或者值得人为加工供人们检索而获得理解。

传统编目程序是记录识别文献的描述信息,对以

纸张为载体的编目是观察其外部并可直接获取,编目的最终目的是为了组织信息、提供检索。对电子出版物的编目仅仅观察外部信息描述是不够的,其内部依然有许多值得和必须采集的信息,现就编目的几个突出问题与大家商榷。

2 电子出版物的分类号著录

电子出版物由于其本身特性即海量的存储和“全盲”的数据信息,可能在电子数据的汇集上出现文献集的跨类,如早期出版的《中国学术期刊(光盘版)》1996年第1期的综合卷,第2期的自然科学与工程技术、社会科学综合卷,第3期的理工专辑、医药卫生专辑、农业专辑、经济法律专辑、社会综合专辑。虽然该光盘设置专辑刊名目录和文章目录是:依据中国图书资料分类法,对入编期刊整刊进行一级分类……对每一篇文章进行二级分类,给出1~3个科目分类号和对应的专辑分类号^[3],在以后的专辑中可以这样处理。但是作为物理整体形式的光盘在著录分类标记符号时就很难进行跨类别标引。随着储存容量的越来越大,这个问题会越来越突出。按照《中国机读目录格式》从675,676,680,686,690和692各种分类字段来看,除686\$a子字段可重复外,其他作为分类号的子字段都是不可重复,只有本字段可重复,如果采用其中某一种分类法,就把光盘数据的自然整体给分割开了,著录分类标记符号容易造成一定的混乱。我们建议在对电子出版物进行分类标引时可重复\$a

子字段,这样就可以解决跨大类别或者跨下位类目的分类问题。

3 电子出版物的内容目录著录

在计算机越来越普及,数字化程度越来越高的情况下,读者对文献检索的需求不再仅仅满足题目、作者、主题词方面传统卡片式的有限检索点,而更注重实际内容如内容目录、提要、文摘、全文等的检索,这些在信息技术发展的今天是非常容易做到的,所以有必要在一旦选择了这些字段就应该把它们作为检索点。按照光盘的工业标准所制造的光盘产品能基本反映主题内容上的区别,所以不必对光盘的编码数据字段和资料特殊细节项作过多的提取,光盘上的目录包括计算机程序名可以增加 329 字段作为内容目录^[4]。为了不与已经形成标准的相关字段发生冲突,我们考虑以增加 329 字段的方法,这个字段本身可以确定为:329 字段指示符 1:内容目录有无检索意义指示符;0 内容目录无意义,主要用于非文字的数据,不便于用作检索点。1 内容目录有意义,内容目录用作检索点,其内容目录是以文字形式著录的,不包括图表。指示符 2:完整程度指示符标识;0 内容目录是不完整的。1 内容目录是完整的。\$a 附注内容,不重复。字段内容注释:这里所指的非文字数据,就目前的数据格式有:.midi(乐器数字接口).bmp(位映图形).tif(矢量图像).jpg(静态图像).gif(活动图像).wav(音波).mov(电影).avi(语音图像)以及.mpg(电影剪辑)这样一些格式,由于技术上的原因要对这些格式的内容进行匹配,涉及到多方面的技术如脉冲信号、数据宽度、像素、颜色、对比度等,除专业方面的要求外,对普通读者来说没有必要进行这样的检索。

需要说明的是 327 字段即内容附注和新增加的 329 字段用作内容目录,在著录时不应该有重复。在《中国机读目录格式使用手册》中的 327 字段的实例列举,实际上仅仅属于内容目录中的栏目而非内容,不作为检索点,因而在没有采用新增加 329 字段前可以根据标准要求著录。如果一旦增加了 329 字段就应该区分这两字段的关系,避免发生冲突,造成数据转换进程的错误。

目录是一个完整的记录,不应按照每一条目内容进行分割而重复,考虑到目录各条之间没有标点符号,生成 MARC 数据后不便于再被准确套录下来,参考《国际标准书目著录》(ISDB)规定的标点符号,使用

在每一条目录之间用逗号、空格(,)进行分割排列,如果是中文内容目录使用双字节逗号(,)进行分割排列,作为 MARC 格式。以计算机软件为例,把可能的软件名包括第一、二层文件夹的名称都进行著录,如果可能也可以对主要文件进行著录。作为检索点,在许多情况下,不是有一个相同的软件名就对用户有用,而是往往需要某个程序文件,通过计算机检索能够查到这个文件比仅仅只检索到软件名的作用也许更大些。

由于各种多媒体教学软件(或称 CAI)是由不同的开发工具制作而成的,完成打包以后就有可能形成各种各样文件组,而文件类型比较复杂,如.exe,.txt,.midi(乐器数字接口).bmp(位映图形)和许多配置.ini(初始化).dll(动态连接库)及其他自定义文件。而 VCD、MTV 光盘目前接触到的主要有一种工具制作,打包后其文件夹由 Cdda,Cdi,Ext,Karaoke,Mpegav,Segment,Vcd 组成,主要是.vcd,.dat 文件。CD 的程序文件比较简单,只有的扩展名为.cda,数据类型是 CD 音轨。这些都是固定的文件扩展名,著录这些文件和大小对文献的管理与检索没有实际的意义,一般不予著录。

采用 329 字段可以有两种方式著录:一种是文本格式,把 CD、MTV 上的曲目用文本方式人工输入,322 制作者附注字段,323 演出者附注字段可以对非演出者和演出者进行著录;另一种是把 MTV、VCD 在播放到目录或主要演职员名时用图像形式录制下来如 MP3 方式,进行连接输入。

如果特别需要对电子出版物中的 CD、VCD、MTV 计算机文件进行著录,应在 230 字段即资料特殊细节项的计算机文件特征 \$a 中,对光盘总的文件大小,各文件数及大小进行著录。一些规范的文件,有其文件属性\详细资料如.avi 有音频格式:PCM,XXXXXHz,8 位,单声道;视频格式:图像大小,数据位,图像个数,传输速率等等。

4 馆藏信息与其他字段的著录

许多“管理集成系统”的软件根据自身传统的馆藏结构,设计 905 馆藏信息字段。目前尚没有一个完全统一的登录号:电子图书及图书附带光盘使用的是图书的登录号,电子期刊及期刊附带光盘使用的是期刊合订本的登录号,或者根本没有号;CD、MTV、VCD、多媒体教学软件则按非书资料直接加流水号;随计算机附带操作系统软件,工具软件,数据库管理

系统,甚至文献数据库等贵重光盘都没能进行正规编目,而是由专人保管。这些分散在各处的电子出版物不利于读者阅览,更不利于建立数据库供网络检索和查询,电子出版物还没有统一可行的规范。对此我们建议单独把电子出版物作为与图书、期刊、资料同类来处理,就可以按照电子出版物的馆藏信息著录,登录号混乱的问题也就自然而然地解决。

特别需要说明的是图书、期刊附带光盘,如何进行著录的问题。应该依据《文献著录总则》规定,分散著录,即另行单独著录。如果继续沿用分室收藏管理,其 905 字段馆藏信息的个别子字段内容不应有重复,否则无法检索到惟一性馆藏信息,给管理造成许多麻烦。

光盘的写入与读出依赖于硬件与软件,随着技术淘汰与更新,直接威胁着前期数据格式光盘的读出。对一般软件光盘的收藏也是必须的,可以根据硬件的需要匹配使用,这样就需要著录光盘使用的硬软件环境。在 3- 附注块的 303 字段即著录信息一般性附注,作为封面或者外套盒“广告语”著录,特别是作为程序软件的安装说明就显得尤为重要,应成为必备字段。否则,一旦硬软件读取环境不再支持普通光盘数据格式,只有考虑馆藏剔旧了。

使用 CNMARC 格式设计数据库管理系统,就是要用已有的数据进行套录,就需要了解哪些检索点是独立的,哪些检索点具有子检索点,这些数据能够供用户使用,所以增加的 329 内容目录字段,对于各种光盘进行内容著录是最主要的出发点。对于从网站上下载的电子数据,应该在 856 电子资源地址与检索字段中的子字段注明来源^[5]。

5 复制品字段和追加数据后的著录

除了由出版发行部门正式出版销售的声像制品、部分小型软件工具、文献数据库和应用系统外,大部分计算机软件、应用数据库管理系统、各种计算机文档及自制数据文件都没有如 ISBN, ISSN, ISRC 号这样的国际标准号。不能认为没有国际标准号的电子出版物是复制品。随着版权的保护,对复制品的管理越来越严格,有些光盘上的文件不能被复制,因而原版电子出版物与复制品之间就存在版权或密钥文件的区别。对复制品的判断可根据光盘的 3 种国际标准类型:只读型光盘(CD-ROM);一次写多次读光盘(WORM, CD-R);可重写光盘(P-C, M-O)来分辨,后两者主要用于复制品制作的光盘,但是有许多文献数

据库出版部门,为了追加数据也使用 CD-R 光盘。如果原版与复制版有区别要在 325 复制品附注字段 \$a 附注内容著录区分细节,或者使用 456 复制为字段作为连接,指示符 1: 填充格;指示符 2, 设为 1 作附注, \$1 子字段内容填被连接的字段数据。

复制光盘在进行第一次编目入藏后,如果认为还有数据因某种原因需要更改或追加,而且该盘还有容量可以继续复制的,应重复 801 记录来源字段指示符 2: 设为 2(更改机构) \$a 国家, \$b 机构, \$c 处理日期。如由北京博利群电子信息有限公司开发制作,电子工业出版社出版发行的《经济学管理学位论文索引》ISBN7-900014-79-9, 每季更新数据,如果进入其网站可以每星期更新,每年又按新数据重新发行光盘,ISBN 号相同,这样在著录 ISBN 号时就有重号。解决的办法:其一把原有的光盘按剔旧处理,重新著录新盘;另一种是在原盘上追加数据,010 字段 \$d 括号内计数据更新费,在 300 字段一般性附注说明追加数据的时间范围及其他相关事宜;再使用 801 字段作数据更改处理,在 905 字段 \$f 增加复本数。

6 非 MARC 字段的使用

虽然 CNMARC 字段已经对电子出版物的信息进行了必要描述,通过检索可以为读者提供有效信息,但是,由于技术上的限制,以及对改动和增加相关字段不能认可,有必要把这些字段设置为非 MARC 字段即不生成 MARC 数据的字段,防止在转换成 CNMARC 数据时,把这些数据一并转出,使人不能理解。另一种确实是要用非 CNMARC 字段,图书、期刊附带光盘一般在 215 \$e 附件中加附注项即 \$e 光盘;或者另行单独著录光盘时加附注项即: \$e 图书(期刊)。但是,读者如何去得到相应的附件?在子字段后没有附件馆藏信息的子字段,需要增加一个子字段 \$f 附件馆藏信息,以指明附件所在位置。

如果对 CD, VCD, MTV 以及有活动图像的多媒体教学软件,在 327 字段即内容附注中著录,由于其 \$a 是可重复子字段可以针对不同的情景进行采集,但是,这些图像数据是根据不同著录者的鉴别能力而采集的,数据量大,而且不好进行规范,所以这些数据也不应生成 MARC 格式进行数据交换。

电子出版物是一个新型的馆藏种类,各种 MARC 格式没有可参照的详细字段,面对越来越多的可存储信息必然造成大量的电子出版物管理上的混乱,有效信息被闲置。沿用传统的图书馆对资料的管理模式

即声像制品来管理电子出版物,数据收集量大,管理不方便而且数据利用率低。把电子出版物从资料中分离出来形成单独的种类进行管理,开展对电子出版物管理、检索的研究以及软件的研制,是对图书馆除图书、期刊、资料外又一品种的有效利用,有利于改进管理模式,进一步加快信息电子化,增加网络信息服务的内容,达到信息资源共享的目的。

参考文献

- 1,4 严彬,王益友,柴晓娟. 电子出版物在数据库中字段选择和 CNMARC 的对应. 国家图书馆学刊, 2000(1)
- 2 严彬,柴晓娟. 图书馆电子出版物的位置所在及其它. 现

代图书情报技术, 2000(1)

- 3 《中国学术期刊(光盘版)》总体介绍. CAJ - CD 导报, 1996-12-01
- 5 朱岩. CN MARC 与全文电子资源检索. 北京图书馆馆刊, 1998(4)

注 本文系江苏省社会发展科技计划项目(BS2000107)的研究成果之一。

严彬 南京邮电学院图书馆信息技术部主任、副研究馆员。通讯地址:江苏南京。邮编 210003。

(来稿时间:2000-08-10)

(上接第 46 页)理使用现有民族古籍整理人员,并采取措施,调动他们的积极性,提高工作效率。(3)抽调一些有条件的人员来做一些民族古籍整理急需的工作。(4)挖掘潜力,扩大培训范围,一些有条件的地方,可办培训班,可以提倡个人带徒弟。

所谓长期目标就是通过民族古籍整理系统正规的教育培训、研究活动和实际工作来培养各层次的专业人员。首先,在文、史方面基础较好的综合性大学里以系、科配合形式建立学科交叉型的民族古籍专业。其次,鉴于民族古籍本身的学科属性,应采取基础宽厚、文理渗透的办法以多学科综合配置的方式设置科目,在此基础上,建立本科生、在职培训和研究生教育同步发展的教育机构。同时,将民族古籍专业建设成为教学与科研相结合的实体。民族古籍专业与民族古籍整理、出版机构挂钩,建立稳固的教学与生产实习基地,用对口设学(科),以求求学的精神去做,达到学用一致。

5.3 开展民族古籍整理研究,建立民族古籍整理学科体系

到目前为止,已有 24 个省、自治区、直辖市相继建立了民族古籍整理出版规划小组、古籍办公室、民族古籍研究所、民族古籍出版社及相应机构。研究内容包括民族古籍本身和如何使民族古籍在社会主义文化事业发展中发挥作用,以及民族古籍的整修、保护、保存的理论、技术和设备。特别是整修保护工作绝不能忽视。因为有许多民族古籍经历了几百年,甚至上千年,由于过去保护不好,老化、脆裂、虫蛀鼠咬,已濒于毁灭的边缘,必须及时抢救整修。

据有关材料介绍,美国国会图书馆通过化学处理,纸张脆裂的问题得到根本解决。其方法是在真空中把书烘干喷以二乙烷基锌,再恢复其温度,可使老化脆裂的书恢复原状,延长保存期 500 年,而且可以不断反复处理。毫无疑问,这些新技术的利用,应该是民族古籍整理研究的一项重要内容。

关于民族古籍的保存,排架也值得研究。过去的传统习惯都是用四合或者六合的布函套,这种办法费工费料,使用起来又很不方便,很有必要进行改进。

民族古籍的科学管理是一门新兴学科,需要研究的课题很多,它应该发展成一门专门的科学,有自己的理论、学科、技术、设备体系。这有待于有志之士继续去开拓创造。

参考文献

- 1 包和平,许斌. 中国民族文献学研究. 北京:中国华侨出版社, 1997
- 2 刘维英. 民族文献检索和利用. 长春:吉林人民出版社, 1999
- 3 李晓菲. 中国民族文献检索. 赤峰:内蒙古科学技术出版社, 1998
- 4 张公瑾等. 民族古籍概况. 北京:民族出版社, 1997
- 5 吴枫. 中国古典文献学. 济南:齐鲁出版社, 1982

注 此文系国家社会科学基金资助课题《中国少数民族文献学概论》阶段性成果之一(批号:00BTQ005)。

包和平 大连民族学院图书馆副研究馆员,副馆长。通讯地址:大连市。邮编 116600。

(来稿时间:2000-06-09)