

严 彬 阚德涛

基于 CNMARC 数据的综合馆藏自动化系统

摘 要 对图书馆馆藏的电子出版物、标准文献、期刊资料等,要建立自动化管理系统,仅依靠 CNMARC 格式,远远不能满足字段需要。必须增加字段,采用 CNMARC 和非 CNMARC 并存的方式,使用有利于满足读者需求的检索项和多媒体阅读工具。参考文献 5。

关键词 综合馆藏 电子出版物 标准文献 期刊 中文机读目录 图书馆自动化

分类号 G254.36

ABSTRACT In this paper, the author thinks that CNMARC format is not sufficient for an automation system for diversified library collections covering all kinds of materials, including electronic publications, standards and periodicals. We should add some fields outside the CNMARC format to fully describe materials to facilitate the access by users. 5 refs.

KEY WORDS Diversified library collections. Electronic publications. Standards. Periodicals. CNMARC. Library automation.

CLASS NUMBER G254.36

大量的新型信息载体文献与传统的纸张载体文献一并入藏图书馆。图书馆的自动化管理系统仅仅对图书期刊类的纸张载体形式进行管理是不够的,仅仅将各种载体文献转化为电子文献在当前也是没有必要的。人们习惯于随时随地、悠然自得地看“书”学习,网络和电子图书的高成本只能使读者坚持这一习惯。为使读者能够尽快检索到所有馆藏并获得所需文献,建设综合馆藏的自动化系统自然是必要的。

1 电子出版物

越来越多的图书馆逐步重视对电子出版物的收藏,并且随着信息技术和存储技术的发展,新型电子信息文献不断出现。不仅有文献数据库如书目、索引、文摘、参考工具书,也有电子期刊、电子图书、多媒体教学课件和计算机软件,更出现像单本电子图书这样可供外借阅读器(e-book),已成为发展最快、覆盖面最广的信息载体。这些已是图书馆增长最快的藏书,并成为主要的标准藏书成分,与图书、期刊、资料并列成独立的种类^[1]。目前电子出版品种多但馆藏管理比较混乱、不规范,主要原因是图书馆对电子出版物的定义范围不明确。根据国家新闻出版署的规定,电子出版物是指以数字代码方式将图文声像等信息编辑加工存储在磁、光、电介质上,通过计算机或者具有类似功能的设备读取使用,用以表达思想、普及知识和积累文化,并可复制发行的大众传播媒体。媒体形态包括软磁盘(FD)、只读光盘(CD-ROM)、交互式光盘(CD-I)、照片光盘(Photo-CD)、高密度只读光盘(DVD-ROM)、集成电路卡(IC-Card)和新闻出版署认定的其他媒体形态^[2]。目前大多数图书馆

缺乏一套有效的管理系统来规范这些电子出版物,把电子出版物仅仅限定在狭窄的范围内,看成是图书种类的附件,和使用非书资料的著录规则,把图书附带光盘、期刊附带光盘、光盘图书或期刊、数据库光盘、多媒体声像光盘、软件光盘等分别用机读和手工方式加馆藏信息归不同阅览室。

《中国机读目录格式》没有涉及到电子出版物,没有中国标准声像制品号(ISRC),没有设定以光盘为载体的声像制品的字段,直到2000年底才要求在现有管理系统上增加ISRC号字段即016子字段,但具体如何著录没有规定。电子出版物的责任说明也不明确。如果按照《中国文献编目规则》对录音资料、影像资料责任说明将作者、设计者、导演、摄影等进行著录就更加“责任说不明”了。电子出版物是以其物理单位(盘、套)为计量的,不是以作品为单位计量的,随着科学技术的发展,存储信息的容量越来越大,可能将大量作品汇集在一个物理单位中。如果使用整件电子出版物的责任说明,在目前的外观信息上很难找到,实际上诸如监制(制作、出品)、编辑(编导、编审)、策划、装帧设计以及导演、摄像、录音等都是整个电子出版物生产过程中的各个环节,不能代表整件电子出版物的责任人。随着对知识产权保护认识的提高,人们承认的是整件电子出版物版权,这在电子出版物外观信息上可以找到,所以将责任说明改为“版权所有说明”是比较明确的,也是比较简单的做法。

电子出版物中的图书、期刊、音像制品在使用国际标准号的同时也使用“中图法”分类号和流水号作为种次号。这种标注的符号由于使用的现行标准并不适用电子出版物,所以不能作为著录依据。GB 3469规定,录音制品载体代码为A,录像制品载体代码为V。分类代码由1~2个字符组成,

根据音像制品的主要学科范畴,按《中国图书馆图书分类法》的基本分类号给出。同样,GB9999-88 规定,一种报刊只能给定一个分类号。GB5795-86 规定,图书分类号由出版社根据图书的学科范畴参照《中国图书馆图书分类法》的基本大类给出。这些标准具体体现在电子出版物的外观信息上,并没有完全得到遵守,不能反映真实的内容,只能作为参考。使用《中国图书馆分类法》对电子出版物进行分类:“G255 各种文献工作,用作各种文献的研究、补充、登记、编目、保管和使用等入此;G225.75 电子出版物”。结合一级类目使用总论复分表-79 非书资料、视听资料,-794 机读资料,也只是电子出版物中的一种分类。除了用“中图法”重复标引,解决跨大类或者下位类目的分类外^[3],很难从综合性内容上获得有用信息。

电子出版物的主要媒体形态在现有的 1—编码信息块中进行数据说明是不适合的。因为编目员无法分辨电子出版物应该归属哪种媒体形态,只能通过内容来确定是按照哪种工业标准制作的,所以采用增加 136 字段用光盘的内容类型进行代码说明,即 136 编码数据字段:电子出版物(光盘),本字段选择使用,字段不可重复。指示符 1 和指示符 2 均为空格(#),子字段 \$a 为电子出版物类型,字符数 2,字符位置 0,1。字符数 1,字符位置 0,数据元名称:1 代表 CD;2 代表 VCD;3 代表信息数据;4 代表多媒体计算机应用软件;5 代表计算机软盘;6 代表文字文本即文献数据库;7 代表其它。字符数 1,字符位置 1,在 1CD 中 0 代

表没有细分类;在 2VCD 中 1 代表 MTV;2 代表 VCD;3 代表 DVD;4 代表其它格式 VCD 如 SVCD、CVD 等;在 3 信息数据中 1 代表图书附带光盘;2 代表期刊附带光盘;在 4 多媒体计算机应用软件中 0 代表没有细分类;在 5 计算机软件中 1 代表操作系统;2 代表程序语言;3 代表数据库管理系统;4 代表实用工具;5 代表应用计算机软件;6 代表综合软件;在 6 文字文本即文献数据库中 1 代表全文;2 代表篇名文摘;3 代表题录;4 代表多媒体;5 代表其它。这样进行内容分类,在多类别综合型文献无法采用“中图法”进行重复分类时,可以使用导航树或某一类别的检索方法,将检索结果显示出来。

322 制作者附注字段(投影和录像资料及录音),如果整件电子出版物的责任说明使用的是版权所有说明,那么在外观信息上能找到诸如监制(制作、出品)、编辑(编导、编审)、策划、装帧设计、以及导演、摄像、录音等责任人,可以选用此字段。323 演出者附注(投影和录像资料及录音)著录主要表演者、演奏者、讲解者、友情出演者等的姓名,以及录音的参加者、演奏者或演奏团体和演奏媒体的名称,作为可检索字段,供普通读者对 CD、VCD、DVD 进行检索。为了避免 327 内容附注字段与增加 329 内容目录字段发生混淆,在 329 字段指示符为 1,即文本数据作检索点时,对计算机软件、CD 进行内容目录检索。210 出版发行等字段也是电子出版物外观信息的主要采集项,用作检索点。具体方法见图 1。

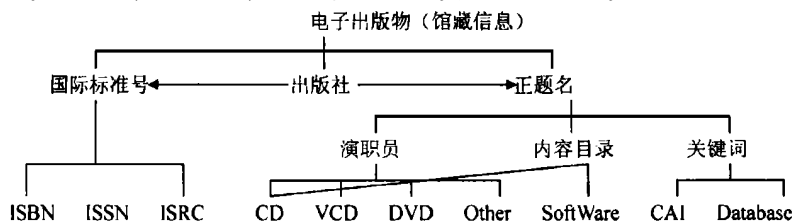


图 1 210 出版发行等字段示意

由于使用了明确的多种途径检索方法,使读者不仅可以找到电子出版物,而且还可以从海量的光盘数据中查询到自己所要的文件。

2 标准文献

标准文献是科技文献的重要组成部分,是直接指导现代化生产建设、科研与管理工作的文件。标准文献越来越引起图书情报单位的重视,设专人专室开展服务。标准文献发展到一定数量后,其数量变化就不大了,只是在内容上进行更新,这一点不同于一般科技文献,而且其类型较少,篇幅较短,使用周期短,一般馆把它作为资料集中管理,很少将数据进入自动化系统。

《中国图书馆分类法》对标准的分类是按大类分在 C6

社会科学参考工具书下的 C65 条例、规程、标准或在 T 工业技术类下的 T-6 参考书, T-65 工业规程与标准, T-652 中国, T-652.1 国家标准,作为综合性标准合订本应分到此类,单本标准按照通常办法是入相关各类,注释中已有说明。《中国机读目录格式》中对 6—主题分析块出现情况说明是供选用字段,但《计算机编目技术手册》附录《字段和子字段一览表》对 690—“中图法”设定为必备字段。这样一来,按照该规定设计的管理系统,编目员不愿意再去使用第二种分类法对标准文献进行分类,这就对分类法的选择上排除了其他分类法的使用。合订本的标准和单本标准都有 ISBN 号,作为图书处理,无论是与其他图书一起排架,还是单独以标准独立排架都会出现架位不整、排架困难等现象,造成实物与数据分离,使读者进行计算机检索

时出现漏检或查阅原文不便。所以许多图书馆把标准集中起来管理,这样就需要考虑数据分类的集中。

《中国机读目录格式》所使用的分类法除“中图法”外,还有《国际十进分类法》(UDC)。该法在 1996 年 4 月前出版的中国国家标准文献中使用,1994 年 7 月以后开始使用《中国标准文献分类法》,1996 年 4 月以后出版的标准还使用《国际标准分类法》(ICS)。但是,《文献目录信息交换用磁带格式》(GB 2901-82)的附录 C《情报检索系统用的数据区内容》(参考件)规定,数据区标识符 086 为本国标准分类号。而 1995 年 12 月出版的《中国机读目录格式》却没有把本国标准分类号作为主要字段列入 6—主题分析块中。实际上,《中国标准文献分类法》在 1989 年 11 月就已经正式出版并使用。CNMARC 格式没有得到及时修订,限制了系统规范化的设计,造成标准文献不能依照标准分类法进行分类和按照人们习惯方式进行管理的问题。

为了有效地发挥标准文献的作用,国家技术监督局组织完成了《中国标准文献分类法》,由 24 大类组成,每个大类有 100 个二级类目。我国现行的国家标准在 1990 年以前使用的是书号,1990 年 2 月以后开始用 ISBN 号。虽然我国在 1983 年编制了《中国标准文献分类法(试行)》,但是直到 1989 年 11 月才正式出版执行,所以在 1989 年 12 月出版的国家标准上增加了《中国标准文献分类法》的分类号。随着标准化事业的发展,各国自行制定的标准分类法已不能满足世界标准化信息交换的需要,为此国际标准化组织(ISO)于 1991 年完成了《国际标准分类法》(International Classification for Standards, ICS)的制定工作,并积极向各成员国推荐,我国于 1996 年 4 月以后出版的国家标准中开始逐步用 ICS 取代 1982 年 6 月正式使用的 UDC,并且继续保留《中国标准文献分类法》。

国家质量技术监督局决定于 1997 年 1 月 1 日起在我国国家标准、行业标准和地方标准上标注国际标准分类号,取代原有的十进分类法分类号^[4]。中文版《国际标准分类法》与英文版 ICS 一样,采用三级分类,第一级由 41 个大类组成,其下分为 405 个二级类目和 884 个三级类目,各级类目之间以实圆点相隔。但是,许多行业标准和地方标准并没有使用《国际标准分类法》,甚至连以前使用过的《中国标准文献分类法》也不用了,如通信行业标准,无法直接利用标准文献上的分类号。

《中国标准文献分类法》公布在先(1989 年),《中国机读目录格式》公布在后(1995 年),应该采用《中国标准文献分类法》对国家标准进行分类著录,而不必再用“中图法”。为了能有效地实现过渡时期的系统改造工作,以便在日后的《中国机读目录格式》修订工作中能够给予考虑,我们提出:(1)鉴于在《中国机读目录格式》中 0—标识块已经定义了 094 为标准号,也就是把它作为文献类型的一种,理应增加《中国标准文献分类法》字段,可以设为 694 字段号,此参考国家标准 GB2901—82 附录 C 的名称排序。(2)通

过“中图法”在 C65 条例、规程、标准和 T-652.1 国家标准与《中国标准文献分类法》进行链接,这样不但可以把标准的合订本和单行本统一归到一个类目下,直接使用标准文献已标注的分类号,而不必再对国家标准按“中图法”进行细分。(3)在系统的检索项中增加《中国标准文献分类法》和《中国标准文献分类法》导航树,便于读者检索。如果系统允许,可增加《国际标准分类法》,与《中国标准文献分类法》并存。

3 连续出版物

虽然 CNMARC 格式规定了 330 字段即提要、文摘或全文,但是与目前大量的商业化文献数据库相比,图书馆花费大量时间来建自己的馆藏全文、篇名数据库有无必要?除了时间上得到保证外,没有任何其他可取之处。如果连续出版物也如同图书那样仅仅让读者检索到有没有某种刊,到没到某种刊,这样的数据对读者来说确实没有多少实际意义。到底什么样的信息既有利于读者,又方便编目和管理呢?这就是目前图书馆自动化系统中对连续出版物设计得不成熟的地方,也是连续出版物没能在图书馆自动化中得到有效采用的主要原因。

由于网络电子期刊资源存在更新滞后或者严重滞后,网络带宽窄,网站管理差,并发用户少而经常登录不上,注册费用高,售后服务差,硬件软件更新快而检索程序开销大,稳定性低等一系列问题,特别是个别利益主体之间为各自经济利益的驱动,没有在满足他人需要的情况下,实现着经济主体的自我利益,因而对现刊的管理是当前图书馆急需解决的重点。

《中国机读目录格式》数据只用于期刊本身,过刊的合订本通过重新编目以图书流通的形式把数据存于系统中,而现刊则是利用该数据进行登到,有的系统是利用 330 字段由期刊管理员将篇名、作者输入,供读者检索。这种方式可以解决文献数据库期刊数据滞后的不足,但是输入的工作量十分大,一般人员很难承受,而且这些数据最多保存一年就需更新。如果使用图像进入数据库管理系统,则数据库管理起来比较困难。采用每期封面和目录页的图像文件,该图像文件不存储在数据库中而是与数据库管理系统分离,可以避免图像文件处理、传输中的问题。另外,图像制作和压缩后应该可以提供篇名和作者的查询功能,尽可能不用或少用了 30 字段,减少题录的选择时间和著录的工作量,避免出现既购买篇名数据库和学术期刊数据库,又自建专业题录数据库所造成资源浪费的现象。

如何著录以期刊为主的连续出版物出现附带光盘?期刊需要装订成合订本后,才能使用 MARC 数据并添加 215 \$e 附件子字段著录光盘及多个光盘馆藏信息;反过来,光盘要等到期刊装订成合订本后才能添加 215 \$e 附件子字段著录期刊及馆藏信息,光盘与期刊始终处于关系

不明的状态。由于使用了电子出版物模块,对每张光盘都贴有条码,可以直接检索到某期的光盘,间接地明确了这种联系。

期刊经过治理后,出现了内部资料准印证号的内部期刊^[5],这种编号由各省级主管部门设计,形式多样没有统一格式。对这些内部期刊的自动化管理,在一般的管理系统中是没有的,《中国机读目录格式》也没有可采用的子字段。要进入管理只有增加非 CNMARC 字段 093 内部资料号,指示符 1 和指示符 2 均为未定义,\$a 内部资料准印证号;\$b 其它资料号;\$c 获得方式和/或定价。

对期刊的责任说明,在使用编目数据套录时经常出现用“该刊编辑部”数据,按《中国文献编目规则》,是负有责任的团体或个人,一般为团体编辑部。如果著录个人则主编经常换有所不当;如果著录编辑部则与 210 字段出版发行 \$c 出版发行者名称重复,检索时会出现漏检或组配无效,实际上第一责任说明应该是主办者,其他责任说明应该是协办者,并非该刊编辑部。

4 资料

资料类型比较复杂,此处仅仅指非书非刊以纸张为载体的文献,一般不作为馆藏固定资产,不进行数据交换,更不可能有很全的外观信息供编目员进行数据著录。如电话簿、列车时刻表、使用手册、会议论文、论坛报告、本校的学位论文等,其特征是没有标准书刊号、出版社、出版地址,更没有版权信息。对这些资料的管理各馆做法不同,随意性较大,多为手工方式管理;在一般的自动化系统中,也完全忽视了对资料类馆藏的管理,成为图书馆计算机管理中的死角。资料的系统管理的主要特征是:(1)不必采用规定的 MARC 数据格式,不需要进行复杂的数据采集,只需提供少量的可供检索字段。(2)在短期内需要进行剔旧,应该有各种不同的剔旧方法,如按时间范围、按某一类、按某一关键词检索的结果选择等;并需要单独建立剔旧库。(3)借阅方式为特短期内,一般在一周以内,需要对不同的资料 and 不同的读者限定不同的出借时间。(4)对其他种类的文献进行整批剔旧后,如果仍然有个别文献需要暂时保留使用,划归资料。

如果将资料数据进入现有的管理系统中,而该系统又没有专门设计资料模块,那么就会造成系统数据的混乱,影响检索速度,统计数据也会不准。这是因为许多系统都是依据 MARC 格式要求进行设计的,具有 MARC 数据的生成功能。除需要增加一些不能为空的馆藏项字段如登录号等外,按照 CNMARC 手册规定,一些字段为必备字段

或子字段,也就是给出了所要求著入数据的最少款项,而系统程序员往往把这些必备字段或子字段定义成不能为空。这些需要著录的数据与资料的实际外观信息有些不符,有些是缺项,使著录工作不能继续进行,只能随便给个字符。资料在馆藏文献中是并行于图书、期刊而独立存在的种类,不能与图书、期刊混淆,其数据仅仅作为本馆自动化系统的一部分,不作交换之用,也就不能用机读目录格式来限定,所以要进入自动化管理就要设计独立的模块。

如果单独设计成资料模块,就要考虑资料与图书、期刊的数据互换:当有些资料需要进行长期保存时,就要考虑将数据转换到图书或期刊模块库中;当有些图书、期刊需要剔旧但有些个别书刊暂时仍有作用,需要将数据转换到资料模块库中。其中涉及到 MARC 数据在各模块中的转移、资料资产的统计、管理规则等的区别。

5 结语

图书馆的馆藏完全依赖于管理系统的设计,依赖于设计者对 CNMARC 格式的理解,同样,一个好的图书馆管理系统按照 CNMARC 格式进行设计后,著录的数据应该按照规定的要求进行采集。设计者要不断掌握图书馆的馆藏结构和种类,制作的管理软件要面向读者,不断进行升级,满足图书馆管理的需要。一个功能强大的、文献类型齐全的、简明方便的图书馆管理集成化系统应该囊括全部馆藏,应该是智能化、多媒体信息的体现,而不是模拟传统文献的管理,以减少体力劳动为目的。

参考文献

- 1 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 电子出版物管理规定.见:中国法律法规大典(光盘).北京:北京大学出版社
- 3 国家质量技术监督局编译.国际标准分类法 ICS(第二版).北京:中国标准出版社,1999
- 4 严彬.电子出版物 CNMARC 著录数据的实际应用.中国图书馆学报,2001,27(1)
- 5 关于期刊业治理工作的通知.见:中国法律法规大典(光盘).北京:北京大学出版社

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。

阚德涛 南京邮电学院图书馆信息技术部助理工程师。通讯地址同上。

(来稿时间:2001-06-25)