

刘 刚

略谈我国大型书目数据中心的建设

ABSTRACT To build a large bibliographic data centre, the unity of MARC forms, the unity of descriptive standards, the unity of classifications and subjects and the use of the standardized file, etc must be strictly carried out and the principles of voluntary participation, mutual benefit, fair competition, etc must be followed

KEY WORDS Bibliographic Data Centre Construction

CLASS NUMBER G356.1

现代科技的飞速发展及在图书馆工作中的应用,已使图书馆发生了巨大变化。作为文献资源电子化重要内容的书目数据建设特别是由众多图书馆参加的大型书目数据中心的建设是促进图书馆现代化的一种重要手段。

1 大型书目数据中心的内容

大型书目数据中心是由众多图书馆组成的。它借助计算机网络系统将存贮在数据库的数据提供给用户使用。在这个数据中心里,要有一个大型的书目数据库。这个数据库的建设要求是:在一个MARC格式下,各文种、各文献类型的书目数据,按照相应的标准、规范进行加工,从而构成一个规模庞大的数据库。

可以看出,大型书目数据中心的数据库就是将现在图书馆中所使用的各种卡片目录的书目信息,按照统一的MARC格式、标准和要求组织起来,提供给人们检索,而不是按照现行的目录组织形式,分别建立各自的书目数据库。它的最大优点是使读者在一个检索环境下,通过一次检索就能得到所需不同文种、不同文献类型的书目信息。这一切在以

往是难以实现的,但在自动化技术和通讯手段发展的今天,这个愿望不仅可以实现,而且为图书馆的发展提供了更广阔的前景。在这方面,国外已有的成功经验可供我们借鉴。

发达国家中,它们书目数据库的建设由于得益于先进的技术支持和其他因素发展很快,在不到30年的时间里已经取得令人瞩目的成就。如北美地区四个较大的书目数据中心RLN、OCLC、W N、UTALS,现在已发展成为以书目数据为主的综合性的信息中心。它们有一个共同的特点,就是都拥有一个大规模或超大规模的书目数据库。在这些数据库中,图书期刊、缩微胶卷等形式的书目数据汇集一起。在这些庞大数据库的基础上,四个中心都开发出相应的功能,给用户提供了丰富的检索途径,可以使用户从不同角度获得数据库中的信息,并通过计算机网络系统进行异地联机编目,从而实现了书目资源的共享,并且为文献资源共享奠定了基础,大大方便了人们的使用。特别是OCLC,在其书目数据库中,目前拥有300多种语言文字的书目数据超过3200万条,馆藏信息超过5亿个,其数据以每年200万条的规模递增。它的英文书目数据覆盖面在95%以上。也就是说,

只有 5% 的英文书目数据需要用户自己制作。这对图书馆的工作无疑产生了重要的影响。实践证明, 一个庞大的书目数据库是它们成功的基础。

2 建设大型书目数据库的意义

2.1 大型书目数据库是现代图书馆工作的基础

在图书馆工作中, 目录是打开这座知识宝库的钥匙, 历来受到人们的高度重视。在传统的手工作业中, 目录卡片的编排一般都是先按照文种, 再按文献形式, 最后按目录组织规则来组织。如图书目录, 先按文种区分为中文图书和外文图书, 再按目录组织规则组织成题名、作者、分类等目录。连续出版物也如是。对某些特殊的文献, 如古籍、地图、盲文读物等, 还要独立组织有关目录。读者要查资料就要按上述目录的组织方法查找, 效率不高是显而易见的。图书馆的内部业务工作中也要组织若干套目录, 以便于采访和编目的查重与加工。为了做好读者咨询工作, 参考咨询人员还要根据需要, 有目的组织专题目录, 以备工作之需, 而这种目录的组织是要经过长期积累和编排才可完成。这些手工作业一直是图书馆工作的基本内容。人们日复一日、年复一年地进行目录卡片的组织工作。当然, 这种工作对于提高图书馆工作的质量, 便于读者查询是不可缺少的。

大型书目数据库的建设, 将把上述工作完全交由计算机来承担。人们坐在电脑终端前就可获得所需要的信息。这个数据库中将不再按文种和文献类型分别建库, 而是在一个 MARC 格式下将它们集中在一起。检索时, 人们可以在瞬间获得各种信息, 有时可得到“山穷水尽疑无路, 柳暗花明又一村”的效果。比如检索一本古籍, 不仅可以了解现存于世的该书古籍版本, 还可以获得后人影印、标注整理的各种版本以及研究、使用它的相关信息。又如查阅一本外文图书, 不仅可以获得

该书的不同中文译本的信息, 还可以得到该书原版的信息。由于书目数据库提供了多途径的检索的功能, 这就大大方便了图书馆参考咨询工作, 工作人员可以根据读者的要求, 利用书名、作者、主题, 特别是逻辑检索的功能, 迅速找到读者所需的信息。这些功能的实现, 无论对于读者还是图书馆工作人员来说都是以往手工作业所望尘莫及的。

2.2 大型书目数据库是资源共享的关键

首先, 在当今信息社会中, 图书馆的收藏量越来越多, 给编目工作带来相当大的压力。图书馆工作人员越来越希望能够借助他人的编目成果来完成自己的工作, 于是就出现了随书配片等工作, 但是效果一直不尽如人意。现在, 这个问题将随着大型书目数据库的建设得到解决。在数据中心里, 只要有一家图书馆输入了某一本书的数据, 那么所有参加馆就可以享受到这个成果, 只需利用这条记录做些与本馆有关的工作, 而不必重新制作。目前, 我国已有不少图书馆使用了北图、申联、深图等机构发行的书目数据软盘, 在一定程度上实现了编目的数据共享。但一个明显的问题就是时效性差, 影响了工作效率。大型书目数据库的参加者可以通过计算机的网络系统瞬时获得数据库中的书目数据, 这就从根本上消除了时效性差的问题。

其次, 在信息社会中, 面对浩如烟海的文献, 图书馆必须做出相应的选择以适应发展。但是, 无论那一家图书馆也不可能全面收藏如此之多的文献。这就要求人们通过采访协调、馆际互借等方式实现文献资源共享。而文献资源共享的前提就是书目信息的共享。在以往, 人们为此编制了各种联合目录, 耗费了大量的人力、时间和经费, 但仍无法满足人们日益增长的需要。突出的问题就是, 手工编制的联合目录不能随着藏书变化及时调整, 制约了人们的需求。

在大型书目数据库中心里, 用户可以随时调用数据库中的数据, 实现了实时服务。人们

在联机编目的同时,即可注明该书目文献的收藏单位。读者在检索书目信息的同时,就能得到该书目数据的馆藏信息,可以就近通过各种方式方便地获得所需资料。并且,由于数据库中数据的馆藏信息可以随时增加或删除,这就解决了传统联合目录中的问题。实际上,一个庞大的大型书目数据中心,同时也是一部由众多图书馆参加的、收藏量巨大的、可随时增删的、电子版的联合目录。这对于资源共享的实现无疑是个关键。

2.3 大型书目数据中心的建设是文献资源电子化的前奏曲

随着科技水平的提高,多媒体和电子出版物会逐渐增多,这将对图书馆产生重要影响。人们预言,这将最终形成“虚拟图书馆”、“电子图书馆”,尽管它们的实现还要经过一段时间,但这是图书馆的一个发展趋势。书目数据的机读化就是这个趋势的前奏曲。书目信息仍是人们获取文献信息的主要途径。并且,人们将来还可以借助大型书目数据中心的网络系统在获得书目信息的同时,获得该书目部分或全部的文献信息。

2.4 大型书目数据中心是图书馆现代化的标志

显而易见,大型书目数据中心的建设将改变传统图书馆的工作程序和方法。它的建设涉及图书馆业务工作的方方面面,将带动这些方面的变革。随着它的建设与使用,目录卡片将从我国图书馆的舞台上逐步消失,读者和图书馆工作人员将从繁琐的查阅、组织目录卡片的环境中解脱出来,人们将从电脑终端上瞬时获得大型书目数据中心中存放的书目信息及相应的电子化的文献信息,这标志着—个旧时代的结束和—个新时代的开始。

3 建设大型书目数据中心的标准

建立大型书目数据中心必须严格执行有关标准化的原则,否则后果不堪设想。在现阶段

与图书馆业务工作相关的标准应该是: MARC 格式的统一; 著录标准的统一; 分类、主题的统一和规范档的使用。

3.1 CNMARC 格式应是我国大型书目数据中心的—标准格式

大型书目数据中心的格式要求必须是一个,这是不言而喻的。因为在现有技术条件下,只有在一个格式下编制书目数据,才有可能存放在一个数据库中提供读者使用。使用 CNMARC 格式作为我国大型书目数据中心的—标准格式是基于以下原因: 首先,该格式是我国图书馆界的行业标准。1986 年北京图书馆自动化发展部提出了《中国机读目录通讯格式》讨论稿,在全国近百个单位征求意见,几经修改,于 1991 年正式出版发行。经过近 10 年的使用得到不断完善。这对于我国的书目数据库建设将起到积极作用。如果对现有标准弃之不用,那就违背了有关标准化工作的原则。其次,在近 10 年中,先后有北京、上海、深圳等地的单位据此研制出相应的软件,并有数家单位利用该格式编制书目数据,并向社会发行书目数据软盘。目前,全国有为数不少的图书馆等单位已经使用了该格式,这为我国大型书目数据中心的建设奠定了基础。第三, CNMARC 格式的基本依据是 UNMARC, 相关的国际、国内标准达数十个。它适用于专著、连续出版物、测绘资料、乐谱、声象类资料等的机读形式书目记录。而 UNIMARC 格式是国际图联为便于世界各国的数据进行交换,在 1977 年制定发布的,后又经过修订,其用意就在于建立国际间统一交换用的机读目录格式。UNMARC 作为—种国际通用格式,要求各国建立专门机构按本国标准编制 MARC 供本国使用,然后将这种 MARC 转换成 UNMARC,从而实现国际机读目录数据库的共享。由此可以看出, CNMARC 格式既适于我国图书馆现存文献的书目数据制作,又便于我们的书目数据在国际间的交流,对于我们的书目数据走向世界

是非常有利的。

对于利用国外编制的书目源数据来加工我们自己的馆藏外文书刊的书目数据, 可以将套录下来的数据通过格式转换的方法, 转换为 CNMARC 格式, 送入我们的大型书目数据中心。

3.2 著录标准的统一

著录标准是图书馆目录制作的基础, 同样也是建设大型书目数据中心的基础。我国的文献标准化工作起步时间虽不长, 但已取得一定成绩。从 1979 年全国文献工作标准化技术委员会成立至今, 已编制 50 余项国家标准。在书刊著录方面, 已经有《普通图书著录规则》《古籍著录规则》《连续出版物著录规则》等的制定和使用。这些标准在大型书目数据中心的建设中也起着举足轻重的作用。

现在所要解决的问题是, 目前国内所制定的文献著录标准基本是用于目录卡片形式的制作, 而机读目录作为机器可以识别的记录还有一些特殊的规定。所以还要解决现有著录标准与 MARC 格式之间的衔接问题, 要根据各文献类型的情况, 制定出适于 MARC 格式的机读目录著录标准。这对于大型书目数据中心的质量建设是至关重要的。

3.3 《中图法》和《汉语主题词表》是大型书目数据中心的标准工具

在大型书目数据中心中, 《中图法》和《汉语主题词表》应是分类和标引的标准工具。目前, 《中图法》的使用范围已遍及全国 96% 以上的图书馆, 而主题词表也将随机读目录的普及而推广。现在向社会发行的书目数据软盘中, 均使用了《中图法》和《汉语主题词表》, 这对于规范书目数据起到了很好的作用。

由于 CNMARC 格式中可以允许同时有数个分类号的存在, 所以, 对于一些特殊文献如古籍的分类, 可以在使用《中图法》的同时, 再标注其他的分类法, 以适应不同方面的需要。而对于转换过来的国外书目数据, 可根据需要, 在保留或删除原有主题和分类的同时,

按照我们的著录要求增加相应的中图分类号和主题词。

3.4 规范档的使用

所谓规范档就是将一个人物或事物的不同称谓, 按照一定的标准归纳到一个标准的题名下, 在书目数据库中, 应该按照这个标准的题名进行检索, 此时可以得到有关该题名的所有信息。规范档一般包括人名、团体/机构名称、地名、丛书名、主题词等。

在数据量较小的书目数据库中, 对规范档的需求还不明显, 但在一个庞大的书目数据库中, 如果没有规范档的使用, 就会造成检索的不准确。如在检索著者时, 就会遇到这样的情况: 对于“墨子”, 是按照“墨子”来检索, 还是用“墨翟”来检索; 检索“唐太宗”, 是查“李世民”, 还是找“唐太宗”等等。这些情况必须由规范档来规定和执行, 才能提高检索质量。

4 建设大型书目数据中心的基本原则

大型书目数据中心与图书馆有着千丝万缕的联系, 它产生于图书馆, 服务于图书馆, 但它有别于图书馆。它应该是一个公益性的文化事业机构, 应参照企业的管理办法实行自负盈亏。国家应该实行一定的优惠政策, 予以保护和鼓励, 使其能够健康发展。数据中心建设的基本原则是: 自愿参加, 互惠互利, 公平竞争。

自愿参加, 即参加一个大型书目数据中心的图书馆应该是自愿的, 而不是靠行政命令。参加者要遵守相应的制度、标准和规范, 不得随意更改, 以保持数据中心业务工作的统一和完整。参加者应打破行业界限, 自由地选择自己认为满意的数据中心。如果用户认为所参加的数据中心不能满足需要, 可以退出转到其他数据中心去。

由于数据中心要自负盈亏, 所以参加者要缴纳相应的费用, 主要是使用数据的费用。

参加者往数据中心的数据库中输入一条数据,数据中心也应支付相应的费用。这样双方互惠互利:数据中心得到的费用,可以用来改善工作条件,更新设备和软件,以便向用户提供更加先进的工作环境;用户在缴纳相应的费用后,就可得到所需的服务和数据,以提高自己的工作数量和质量。

我国各地发展不平衡,幅员辽阔,所以至少应该有二至三家大型书目数据中心,以适应各地的要求。几家数据中心的同时存在,应打破行业垄断和地区垄断,促使它们之间的相互公平竞争,借此加速我国图书馆现代化的建设。

由于用户可以自由选择数据中心,各数据中心提供的条件和服务手段是否能满足用户的需要就成为吸引用户的关键。各数据中心必须改善自身条件,积极提供用户需要的服务,吸引用户,从而扩大、增强自己的实力。如此周而复始,在发展中逐渐壮大。可以说,大型书目数据中心的建设必将推动我国图书馆工作的现代化进程。

刘刚 北京图书馆自动化发展部副主任。通讯地址:北京白石桥路39号。邮编100081。

(来稿时间:1995.12.21。编发者:黄文田)

(上接第61页) 过鉴定,1996年内由北京大学出版社出版。南京图书馆、南大图书馆等正在联合编制《中国社会科学引文索引》。这两部大型工具书将具有填补空白之功用。1992年出版的《中文核心期刊要目总览》正在修订,也将在1996年出版。全国高校图书馆期刊工作研究会正计划对国际科技核心期刊目录进行修订并准备编制专题性核心期刊表、中文评论文摘和索引等检索工具。联合攻关有固定的编辑班子和连续修订制度,又有全国性权威专家审核,因此可以坚信,我国中外文各专业核心期刊的筛选必将逐步完善。

在我国,核心期刊研究正处于发展之中。我们既要大胆探索,又要谨慎求证;对于核心期刊的作用,既不能一味拔高,也不能一概否定。这才是我国核心期刊研究应走的正道。

参考文献

- 1 Broadus R. N. Selecting Materials for Libraries 2nd ed. New York: H. W. Wilson, 1973: 101
- 2 y-k. Chung (U SA). Bradford Distribution and Core Authors in Classification Systems Literature. *Scientometrics*, 1994, 29(2)
- 3 Michalos A.C. (Canada). Prices and Impact Factors of Interdisciplinary Social Science Journals

Scientometrics, 1995, 32(3)

- 4 Coleman S R (U SA). Disciplinary Variables that Affect the shape of Bradford's Bibliograph. *Scientometrics*, 1994, 29(1)
- 5 Williamson E. Operations Research Journals: An Annotated Core List. *Serials Review*, 1992, 18(4)
- 6 Nisonger T. E. A Ranking of Political Science Journals Based on Citation Data. *Serials Review*, 1993, 19(4)
- 7 徐雁《文献情报学丛书》组稿出版. *大学图书馆学报*, 1992(6)
- 8 包忠文《核心期刊概论》序. *江苏图书馆学报*, 1996(1)
- 9 江乃武等. 连续出版物管理研究进展. 上海: 世界图书出版公司, 1993
- 10 杨弘远. 关于我国自然科学期刊工作的思考与建议. *中国科学报*, 1996-03-04(2)
- 11 刘维义. 文摘法与引文法测单学科核心期刊的本质区别. *中国图书馆学报*, 1996, 22(1): 11~15, 43

叶继元 南京大学图书馆研究馆员。通讯地址:南京市汉口路22号。邮编:210093。

(来稿时间:1996.4.1。编发者:丘峰)