中国图书馆学报(双月刊)1999 年第 2 期
ZHONGGUO TUSHUGUANXUEBAO

●王运堂 李勇慧

关于善本古籍书目数据库建设的回顾与思考

ABSTRACT The construction of databases of rare books has just begun in China. Based on the experience of the construction of a database of rare books in Shandong Province, the authors propose some methods and pose some open problems.

KEY WORDS Database of rare books. Database development. Methods.

CLASS NUMBER G258.94

具有5000年文明历史的中华民族留下了近10万种古籍,其丰富深厚的文化内涵是我们研究传统文化、弘扬民族精神取之不尽的宝库。而古籍书目是古籍整理工作者走进这座智慧宝库的金钥匙和领路人。历代公私藏书家和学者都把编制和阅读目录当作治学的津梁。为此,近现代不少图书馆不但编制自己馆藏的卡片目录和书本目录,还相互协作统一编制一些联合目录,如《中国地方志联合目录》、《中国丛书综录》、《中国古籍善本书目》、《中国古籍总目》等。这些古籍书目的出现,为学术界提供了极大的便利。但要使学者传统治学方法"皓首穷经"的局面得到彻底改变,使人们可以在最短的时间内查遍所有有关古籍目录及后人研究成果,还需利用科学技术尤其是计算机在资料的储存、整理、检索、数据的统计以及索引的编制等方面的优越性。尽早建立古籍书目数据库,才能真正做到资源共享,为人类文明的发展与进步做出贡献。

1996年,面对新信息技术提出的挑战,山东省图书馆(以下简称省馆)开始了古籍书目数据库的建设工作。

1 建古籍书目数据库应具备的条件

目前国内中文新书、报刊等书目数据库的建设可谓蓬蓬勃勃,而古籍书目数据库却由于各方面的限制举步维艰。山东省馆有90年建馆历史,古籍藏量几达80万册,列全国第7位,因此对如何建立古籍数据库非常慎重。首先派人对全国各大图书馆古籍部进行详细咨询或实地考察,并多次与国务院古籍整理规划小组联系。通过多方考察和论证,认为建古籍书目数据库需要具备5个前提条件。

1.1 与国际接轨的国家古籍著录标准

统一的、行之有效的国家著录标准,是建立书目数据库首先要解决的问题。山东省馆古籍编目自1987年开始使用GB3792.7-87《中华人民共和国国家标准·古籍著录规则》。虽然这一规则的颁布使国内传统著录有了统一的可执行的标准,但这只是为手工著录而编订,不符合国际通用的能让计算机识别的著录规则。因此在古籍书目数据库中,文献检索标目必须规范,尤其是题名和责任者名称及古籍中不同类型的版本形式应予以规范控制,并确定不同名称的参照关系。1996年10月,由中国文献编目规则编撰小组编撰,全国情报文献工作标准化技术委员会及中国图书馆学会推荐使用的《中国文献编目规则》出版,较好地解决了这一问题。"此标准是为适应国际文献工作一体化的发展趋势,顺利实现中外文献书目

信息交流，坚持以《国际标准书目著录》(ISBD)和中国文献著录国家标准为依据，并参考《英美编目规则第2版》(AACR II)1988年修订版，同时，充分考虑中国文献语言和书目传统特点。"

1.2 与国际接轨的国家标准机读目录格式

统一的、共同遵循和使用的数据库形式，是建立规范的、成功的数据库的必要前提之一，也是数据库的生命所在。如果没有标准目录格式，数据库就无法进行交换，无法真正实现资源共享。目前，国际图联开发的UNIMARC是世界各国图书馆通用的机读目录格式，各国都根据它做了适合本国国情的修改。1996年，我国文化部也颁布了《中华人民共和国文化行业标准·中国机读目录格式》即CNMARC，这个标准亦是"等效采用了国际图书馆联合会(IFLA)的UNIMARC，它是供中国国家书目机构同其它国家书目机构之间以及中国国内图书情报部门之间，以标准的计算机可读形式交换书目信息，在数据规范方面为书目数据库的建立和书目数据处理提供参考和依照。""同时针对中国出版物的一些特殊情况和中国机读编目的实际做了必要的扩充。"如在记录头标区加"u=拓片"、"V=善本书"等。

1.3 统一的分类法

若要编制规范标准的数据库，一定要有公认的、统一遵循的类表，而且类目名称也确定。中国古籍使用的分类法，在1911年前，虽然也呈现着不同的状况，但基本上还是以四部法为主。辛亥革命后，类分古籍虽然仍以四部法为主，但一些新的分类方法相继产生。如山东省馆在50年代末普通古籍用自订的"十分法"。虽说现在古籍分类没有像中文新书分类那样有统一的《中图法》，但自1979年开始编撰的《中国古籍善本书目》和现在正在编制中的《中国古籍总目》都用四部法。各大图书馆古籍部也大都采用四部法类分古籍。四部法可成为类分古籍的一个公认并通行的分类法。

1.4 适合古籍特点的汉字平台

古籍书除在装订形式上有别于新书外，其最大的特点就是使用繁体字乃至甲骨、金文、篆、隶等带有典型时代特征的古文字字体。在古籍数据库的建设中一定要先考虑这个因素。目前，国内使用的汉字字库有GB2312－80(6753字)、台湾的BIG5码(13053字)，ISO10646的包括中、日、韩三国用字的大字符集(20902字)。GB2312－80的6753字容量太小；台湾的BIG5码其汉字内码与内地不兼容；ISO10646没有合适的录入方法。只有1995年国务院古籍整理规划小组组织人力开发的"四库大汉字平台"即SKDOS，是建立在UCDOS3.1平台上的中文汉字深层应用平台，完全兼容其他中西文系统建立的文件，对于以前用其他系统建立的文稿、数据库，可以不加转换地用于此系统。并可与BIG5码、ISO10646码、方正系统双向转换，也可供五笔、拼音等输入。

1.5 通用编目软件

上述条件都具备以后，如何能找到一个既能包容古籍编目的所有著录款项，并能具有实现各种情况的资料统计、自动排序等功能的编目软件，就成了一个至关重要的问题。经过多方考察和论证，山东省馆最后决定选择北京息洋电子信息技术研究所开发的GLAS集成系统软件中的GCS编目子系统，主要是因为它与四库大汉字平台是兼容的。

2 培训队伍

图书馆界现有的古籍整理队伍人员很少，正处于青黄不接时期，但有一批30岁左右的年轻人，他们稍通古籍，又容易接受新知识。就山东省馆请有关方面的专家对古籍部编目人员进行集中强化培训。一是为保证建库质量，二是给图书馆古籍界培养一批掌握新知识、能熟练运用计算机的新型人才。具体做法是：

2.1 学习《中国文献著录规则·古籍著录规则》

这次要建的善本书目数据库是回溯建库。就是将手工环节积累的全部馆藏业务目录卡片以MARC格式录入中央库，"对号入座"。在

学习中发现，新旧编目规则的最大不同之处在版本项。新编目规则充分考虑到机读目录的发展要求，对著录项目职能的区分较为明确，利于机器的识别和阅读。如张金吾《爱日精庐藏书志三十六卷》(嘉庆本)，在旧编目规则作"清嘉庆24年(1819)海虞张金吾爱日精庐木活字本"只能有一个检索点；而在新编目规则中，为规范控制，加强数据库的检索功能而做成"木活字本·海虞：张金吾　爱日精庐，清嘉庆24年(1819)"。这样"木活字本"、"海虞"、"张金吾　爱日精庐"、"清嘉庆24年"可做4个检索点。

2.2　进行计算机知识及输入方法培训

目前国内图书馆使用的编目软件大都是基于DOS基础上开发的，包括1996年版的GCS也不例外，它对单用户版的实用要求是DOS5.0以上版本。因此，请计算机室的专业人员重点讲DOS操作系统的基本常识和操作方法，并在输入方法中重点抓了五笔字型输入法的培训。

2.3　进行CNMARC培训

由参加过全国CNMARC培训班的同志进行串讲。对较难掌握的"记录头标区"、100"一般处理数据"及105"图书编码数据"，除重点讲解外，还把这3个字段所用的数据元素代码制成表格贴在计算机前，供录入工作人员随时参照使用，收到了良好的效果。

2.4　重点学习GCS编目软件

由北京息洋电子信息研究所开发的GCS编目子系统，其优点在于操作灵活方便，检索点可按自己的需要任意设置。但它是属于全屏幕编辑方式，对还不很熟悉CNMARC的人很不方便。为此山东省馆又编辑了"GCS编目系统使用手册"，将经常使用的字段内容浓缩成一张表贴于计算机旁，加快了建库的速度。

3　操作实施

经过以上的调查论证及各项培训后，自1996年9月山东省馆善本古籍书目的回溯建库工作正式开始。建库的宗旨是，数据库的设置完全按《中国文献编目规则》和CNMARC来执行。因为随着科技的进步，编目软件的升级及新汉字平台的出现都是势在必行，只要保证数据的准确性和标准化，就能保证这个数据库的长久生命力。另外，在建库中也充分考虑到古籍的特点，完全保留原著录中的繁体字、冷僻字、异体字、避讳字，其他附注字段也完全按繁体字录入，充分体现了古籍的原貌及古籍数据库的时代感。而且在制定回溯策略时，将馆藏古籍目录卡片分为5部分：善本目录、海源阁目录、易庐目录、地方文献目录、普通线装古籍目录。其中标准数据最多的是善本目录、海源阁目录、易庐目录。考虑到善本书的重要性及利用率的特点，最后决定先对善本目录进行回溯建库。因为善本目录经过了70年代末80年代初《中国古籍善本总目》时的统编，著录的准确性高，且描述文献实体的规范记录多。下面详细介绍操作步骤。

3.1　做缺省工作单

用息洋GCS编目软件做古籍书目数据库山东省馆是第一家，没有任何可借鉴的经验，因此在建库初始没有大规模展开，而是由搞计算机的和搞古籍的各1人先动手试验。经过讨论，确定各著录款目的相应位置。为便于操作，在原软件"缺省工作单"的基础上，结合古籍著录的特点，把常用的字段及出现频率较高的著录词都做到工作单上，"ctrl + y"键存入普通库的缺省工作单文件中，以提高工作效率。此工作单之所以做两个"306▼a"字段，是因为古籍著录中"出版发行项"附注特别多，如行款、序跋、牌记、印章、避讳、作伪、挖改、刻工姓名等，而这些又都是研究古籍版本的重要依据。而软件设计只有在紧接"▼a"后才能检索到，所以如印章、刻工姓名等最好与其他的附注项区别开，单独在"306▼a"后分别著录，加强数据库的检索功能。

3.2　设检索点

数据库最大的功用在于检索。数据库中检索点越多越便于使用，质量也就越高。根据

古籍著录本身附注项繁杂等情况，设置了近20处检索点。如：010▼b装订形式；200▼a题名；205▼a版本；210▼a刻书地；210▼c刻书者；210▼d刻书时间；225▼a丛书零种所属丛书名；327▼a丛书名；304▼a题名与责任者附注；305▼a版本与书目史附注（如"入四库存目"等）；306▼a出版发行项附注（如印章、刻工姓名等）；312▼c相关题名附注；610▼a非控制题词（如"山东地方文献"）；701▼a人名等同，即第一著者；701▼f著者时代；702▼a次人名等同，即第二著者；702▼f第二著者时代；905▼c索书号，如905▼c善1233③（上角码③表示善本）。这些检索点都是由有关专人设定，其他人不得擅动软件的程序。经过特藏部同仁1年多的努力，至1998年4月，全部完成善本目录5328种及数千条子目的回溯建库工作。

4 古籍书目数据库建设中应解决的问题

经过实践，古籍书目数据库的建设还存在4个方面的问题。

4.1 支持古籍整理研究的汉字平台

目前所用SKDOS，用全拼及五笔输入方法只收录了国家标准中的一级和二级汉字6763个，只能说基本上解决了汉字字数的问题，但输入输出的精度不高。只有区位输入法采用的区位码才能把国际标准化委员会通过的ISO10646标准的20902字全部编码，所以，汉字的不足严重地影响了输入输出的速度。还有些馆所用的是郑码，郑码虽是一种字根通用码，但重码仍比较多，要达到"盲打"的效果，还有一定差距。因此，由权威确定并通过国家标准化委员会颁布为标准的基本汉字字库的确定及汉字平台的开发是当前亟待解决的问题。

4.2 能包容古籍编目所有著录款项的通用编目软件

应尽快结束古籍编目软件春秋争霸的局面，由国家权威部门组织一批既有古籍知识又通计算机的人才，编制一个通用的古籍编目软件，为数据联网及交换提供一个良好的环境。该软件要注意加强检索功能及二次开发功能。如可用已建的数据库转换成能自动排序并包容所有附注款项的书本式目录。

4.3 将四部分类法数码化

四部法在我国历经千余年。作为一种千古流传的分类法，在CNMARC中应有它的一席之地。另外，应有国家古籍整理权威部门组织人力用一种数码的形式取代四部法，如同《中图法》一样。使数据库可用更少的空间储存更多的数据，也达到便于检索的目的。

4.4 CNMARC中的记录类型有待改进

在CNMARC"记录头标区"的执行代码中，记录类型"v=善本书"是中国根据自身的情况增加的。但记录类型中同时又有"a=文字资料印刷品"、"b=文字资料手稿"，而记录类型在CNMARC的规定中只能占一个字符位。如一部属于善本的手稿本，在"记录类型"中或只能用代码b或只能选择v，不能两全，这是CNMARC应解决的问题。

参考文献

1 中国文献编目规则编撰小组.《中国文献编目规则》前言.广州：广东人民出版社，1996
2 中华人民共和国文化部.《中国机读目录格式》前言.文化部编印，1996

王运堂 山东省图书馆馆长。通讯地址：山东济南市。邮编250011。
李勇慧 山东省图书馆。通讯地址同上。

（来稿时间：1998-11-09。编发者：翟凤岐）