第11卷第2期 广西科学院学报 Vol. 11, No. 2
1995年6月 Journal of the Guangxi Academy of Sciences Jun., 1995

机读目录数据库建设
——试论网络化资源共享的前期工程

Construction the data base of machine reading Catalog
—Discuss the earlier stage engineering of network resources sharing

陈登岳
Cheng Dengyue

（广西师范大学计算分析测试中心 桂林 541004）
(Computer Analysis Testing Center Guangxi Normal University. Guilin, 541004)

A
摘要 书目数据的机读化是图书馆实现自动化的基础工程，而机读目录数据的标准化，将直接影响图书馆资源共享的实现，格式标准和著录标准是两种相关又独立的标准。图书馆自动化过程中的数据标准化，主要以著录标准化为重点。书目数据机读化可归结为人工输入与套录两种方式，分别讨论了各自做法及要点。为保证数据质量，每道工序要建立质量反馈控制体系，并分析了影响数据质量的两道关键工序：文献标引和数据录入。

关键词 机读目录 标准化 质量控制 数据库 图书馆，文献资源共享，

Abstract Data machine reading is a basic engineering of library automation. The standardization of Catalog data machine reading will affect the implement of library resources sharing directly. Format standard and register standard are two both dependent and independent standards. Register standard is a main standard in library automation process. This paper introduced two aproaches of data machine reading. Quality feedback control system in each part of process must be established in order to ensure the data quality. Meanwhile the two main processes, literature citation and data input, were also analysed.

Key words machine reading catolog, standardization, quality control

现代科学技术的发展，使人们认识到计算机网络技术和远程通讯技术，将对人类科学和国民经济所产生的深远影响。近年来各种教育科研的局域网络，纷纷出现，在国家教委支持下，分布于全国高等院校的计算机网络，迅猛发展，可以期待，在不久的将来，一个全国高校联机网络将出现在华夏大地上。

由计算机网络给人们提供的文献资源共享，影响最为广泛，受益最为深刻的首推高校图书情报系统。高校图书情报系统，除了图书馆这个文献信息中心以外，还包括各院（系）资

1995-05-10 收稿。

料室、情报室以及档案室、科研处、教务处、出版社、电教中心、学报编辑部等具有情报职能的机构。它是学校科学决策、民主管理的支持系统，也是教学、科研和科技开发的信息保障和支持系统。这是一个多层次的信息情报体系，整个体系实现网络化资源共享，是一个巨大的系统工程。图书馆自动化系统的建立，则是整个系统工程中一个重要的子系统。图书馆的自动化系统，包括图书编目，图书采访，索引编制，流通管理，情报检索等诸子系统。其中实现自动化管理的核心工程，是机读目录数据库的建设。它是网络化资源共享的前期工程，又是一项工作量巨大又难于实现的基础工程。

1　图书馆机读目录的标准化建设

机读目录是机器可读目录的简称，是文献目录载体发展到本世纪60年代出现的一种新型目录。目录载体随着存贮介质的改变而不断发展，由书本型、卡片型、穿孔卡片型、缩微型发展到今天的机读型。机读目录是供计算机阅读的一种目录，其文献目录信息被记录在计算机存贮载体上。即计算机磁盘、磁带、光盘上。机读目录产生于美国国会图书馆，英文Machine Readable Catalog，缩写MARC。MARC是专指美国国会图书馆的机读目录，也称为LC—MARC。

70年代以来，美、英、法、德及其他一些国家也纷纷进行研究、规划、试验和建立自己的机读目录系统，机读目录在世界上得到广泛的应用。在实践中，人们认识到，必须要有标准格式记录的标准著录才能有效地交换信息，也就是说，在计算机网络条件下，彼此交换编目数据，必须要以一定格式和方法为标准，计算机才能识别。国际图联（IFLA）1977年推出了采用通用格式结构的国际书目数据通信格式，简称UNIMARC，以后又几经修改与增补，UNIMARC已在全世界书目记录格式标准化中发挥了重要作用。

中国机读目录格式（China MARC Format）简称CN—MARC，是以UNIMARC为基础的规范数据交换的国家标准。1982年2月国家标准总局批准，定为正式国家标准（GB2901—81），北京图书馆1986年开始编辑《中国机读目录通讯格式》，几经修改，1991年2月由书目文献出版社正式出版，从此，中国便有了自己的标准机读目录格式，为我国图书情报处理自动化、网络化奠定了基础。

中国机读目录通讯格式（CN—MARC）在书目数据标准化上，仅提供了标准格式，其内容表现形式，即著录标准，则是数据标准化的重点。对于一个图书馆自动化系统而言，标准化实质就是要求系统具备完整齐全正确的数据项目。标准化数据产生，要依赖于著录标准和格式标准。标准格式要以标准著录为基础，标准著录要由标准格式来表达，在计算机网络化的环境下，标准化数据才能实施资源共享。

中国机读目录有一百多种数据字段，在数据库建库过程中，要严格执行国家有关标准，制订出适用于本系统建库工作的具有一定权威性的规范及规定，这样才能满足数据“准确性、一致性、完整性、规范性、适用性”等多方面质量要求。建库中所进行的标准化工作包括下列内容：

（1）数据输入工作单　根据本系统对数据的要求，依照《中国机读目录通讯格式》（CN—MARC）选择必要的数据字段，设计出适合本部门使用的”数据输入工作单”。工作单一旦制定，就要维护其稳定性，否则会造成严重数据混乱。

（2）图书分类　按中图法或沿用本系统分类体系。图书分类是按书的内容学科属性来系

统地揭示内容特征的方法，书目分类号可以提供族性检索，因而大部分读者多从这条途径检索自己所需要的文献。高质量的分类标引工作，应达到准确性和一致性（表现在同类书，同种书的各种版本，同种书的各分卷册都能集中在一起）。

（3）主题标引　主题标引是利用规范化的语词系统（如：《汉语主题词表》等）作主题标识，再按一定的标引规则和方法，把图书中论述的主要问题的自然语言，上升为规范的专指词。在图书情报自动化管理中，更能表现出主题检索的优越性。主题词标引，可以参考GB3860—83《文献主题标引规则》。

（4）文献著录　著录数据必须达到两点要求：(1) 标准著录，中文图书依照GB3792.1—83《文献著录总则》，GB3792.2—85《普通图书著录规则》，对文献内容和形式特征进行分析、选择和记录。(2) 著录项目要齐全。依照本系统的“数据输入工作单”，将每篇文献的有关著录项目（机读目录中字段和子字段）详细准确著录下来，成为一条完整的记录输入计算机内。

（5）代码代号　与中文图书建库有关的代码代号标准有：

GB 5195—86《国际标准书号》

GB 3304—86《世界各国和地区名称代码》

GB 3469—83《文献类型与文献载体代码》

GB 4880—85《世界语种代码》

GB 4881—85《中国语种代码》

（6）规范文档　对团体名称，个人责任者及外国责任者的译名等，要建立必要的规范文档，使其数据标准化、规范化，这些文档的约束作用，将对数据一致性产生深远影响。

2　机读目录数据库建立途径与方法

图书馆实现自动化最为首要的任务就是建立机读目录。书目数据转换成机读目录是一项既费时又费力的工作，一旦转换成机读目录后，就可以以磁带、磁盘、光盘的形式提供给其他图书情报部门，共享编目成果。数据一次录入，可以多次调用。因此建立图书馆书目数据库的模式，可以是手工录入和套录两种。

2.1　手工录入

对于机读目录数据录入人员的专业培训，一是要熟练掌握计算机操作程序和机读目录数据格式；二是必须掌握基本编目技术，熟悉卡片目录的著录规则。

编目人员直接上机编目，输入数据，要优于专职录入人员。编目人员不仅熟悉著录规则，而且熟悉所著录图书的外表特征，也熟悉各种文献特征的不同描述和表述，由编目人员输入数据，易于保证数据的标准性、完整性、一致性和准确性。

人工输入计算机编目可分以下几步进行：

（1）收集书目数据，分析和著录　计算机输入数据，是以卡片目录为依据。由于手工操作原因，目录卡片或多或少都存在一些问题。例如：图书的外表特征与卡片著录是否完全相符，各种款目的著录项目是否齐全、正确。要按照本系统的要求，补齐全部著录款目、数据。把卡片目录回溯转成机读目录时，这是首要工作。

（2）手工编辑处理　在卡片上对各著录事项与各字段、子字段填加标识符、分隔符及编目处理需要的有关代码。

（3）将编辑处理后的书目数据输入计算机。

（4）计算机处理　将输入的书目数据经过编目程序处理后，形成机读目录的格式结构。

（5）存贮　将标准机读目录存贮在外部介质上。如磁带、磁盘、光盘等，成为可交换的计算机目录载体，或作为数据库中的共享数据源。

（6）从计算机中输出分类、主题、责任者等各种形式的目录。

2.2　套录

计算机编目方法，任务重，工作量大。随着国内图书馆自动化工作的发展，出现了一系列标准书目数据源，即商品化机读目录。如：北京图书馆的中文图书机读目录，中科院文献情报中心的西文期刊联合目录，上海图书馆的中文期刊目录等，都已发行了磁带版和磁盘版机读数据库。由于光盘技术出现，近年来有各种书目型数据库CD—ROM产品面世，同时配有相应检索软件，可输出符合ISO—2709或UNIMARC格式书目记录。由于国内出现的丰富的标准化机读目录数据；引用标准化数据源，可以帮助建立本馆书目数据库，这是一种快捷的方法和途径。即被称之为套录的方法。其特点是所提取的书目记录项目齐全，著录标准化、节约了人工输入数据的大量人力和时间，使图书馆自动化早日实现。

套录可分以下几道步骤：

（1）阅读外来机读目录　将磁带、磁盘等载体读入计算机。若没有检索软件，还要编制相应软件，打印出全部书目数据，成为可阅读的文字目录，对书目数据进行选择。

（2）检索记录　为了便于书目查询和检索，收到机读目录后，应先作试检索。首先要明确记录格式，可检字段，所用分类法，主题词表以及机读目录中收录图书的范围等。试检索，可用标准书号、作者、书名等作为检索项目。必要时，也可使用全部项目检索，得出可靠的检索数据，以掌握机读目录真实情况，为套录做好准备。

（3）修改记录　利用标准数据的机读目录，一般可不作修改，只需检索和套录。为了更有效地利用机读目录，对目录数据还要作适当修改，以满足本馆需要。例如：增加本馆馆藏号，分类号或联合目录号等，使之成为实用的编目成果。

（4）文档管理　如果要累积机读目录数据，就必须进行文档管理。文档管理分两种：一种是对外来机读目录不作任何修改的文档管理，这种管理较简单，一般是将机读目录载体按分类或其它标识编列，保存在相适应的环境中，随时上机使用。另一种是将外来机读目录数据与本馆机读目录数据合并后，供使用。外来机读目录一般都定期生产，周期性提供，所收录数据范围较广，种类较多，使用不便。可将同一类型书目数据转录到同一磁带或磁盘上，把综合性变为学科性，按学科汇集记录，成为更适用的机读目录。

3　数据加工过程质量控制

书目数据是数据库的“灵魂”，数据质量直接关系到数据库的存在价值。

数据库制作过程由若干工序组成，每一道工序都要建立质量反馈控制体系。影响数据质量的主要关键工序为文献标引和数据录入。

文献标引，一是文献外表特征的著录，如题名、责任者等；二是文献内容特征的著录，主要是分类标引和主题标引。分类标引在传统的书目卡片加工中，人们已积累了丰富经验。图书馆自动化管理中，主题检索是一个必不可少的功能。因此，主题标引就给传统的手工编目，增加了一个新内容。在对文献进行主题分析的过程中，除需确定主题类型与结构外，尚需剖

（下转第70页）

题，这并不是什么奇怪的事情。但是目前还有相当一部分用户对此不理解，以为是负担，其实是对这一问题认识不够。前面我们提到过应用软件的开发过程，其过程大至上可以划分成：用户需求→系统分析→系统设计→编制程序→投入使用。如果把投入运行以前称为第一阶段，投入使用为第二阶段，那么前者对于后者来说是静态的，但大部分用户的需求是动态的。解决这对矛盾的根本途径就是对现有的应用软件进行维护和保养。例如，由于工作（或应用）环境的改变，应用软件也要在功能上作出相应的修改。还有，由于信息量的增大，需要对原有应用软件的性能加以提高。再有，或者是应用软件自身的隐含错误，有可能在运行相当一段时间后才被发现，对这种情况要作排错修正处理。因此，从某种意义上来说，应用软件的使用过程，也就是它的维护、保养过程。既然存在维护、保养过程，在作项目规划时就应该考虑这方面的经费。

4　结束语

使用办公室自动化系统不是目的，只是一种手段，一种用来解放生产力和提高办公效率的工具。在本质上与过去旧有的人工系统中的手工操作没有太大的区别。我们要正视OAS的发展，要看到它可能发挥的功能和效率。但是，这并不等于把它神秘化，也并不等于对它寄予不相称和不切合实际的期望。要面对现实，脚踏实地的去做这项工作。从一件简单的工具到一套复杂的系统，能否充分发挥它们的效率，关键还是人，以及这些人所处的环境。因此，我们应该把看问题的焦点从物移到人，因为我们要用OAS去解决人的问题，而不是用人去解决OAS的问题（开发期间除外，但目的是一样的），如果本末倒置地处理问题，其结果也将是相反的。

（上接第64页）

析主题的中心部分、动态部分和限定修饰部分，以便对文献内容所涉及的主题概念进行精选与取舍。在精选主题概念时，标引人员应充分考虑读者检索要求，分析选定对读者有实际意义的主题概念。应充分考虑主题分析的全面性、专指性，最大限度地满足查全、查准的需要。选定的主题词，必须是《汉语主题词表》使用的主题词，非正式主题词不能作为标引词使用。由于《汉语主题词表》对新学科、新技术的局限性，可以考虑“主题词＋关键词”的标引方式。

文献加工人员，在文献分类、主题标引后，直接上机编目。熟练地掌握各项著录规则及机读目录通讯格式，成为数据录入质量保证的基本条件。

录入数据在进入总库前必须进行校对。为了保证质量，可以设置三道关卡：一是由计算机本身程序来解决，有的数据可由机器校验。二是编制校对软件，由计算机校对。三是由编目专家最后把关。必须把数据错误消灭在进入总库之前，保证书目数据准确无误。

机读目录数据库的建设，为网络化资源共享，提供了基本条件。建立机读目录数据库，在高校内部实现文献情报网络化，能为高校间联网，或加入更广泛的国内、国际情报网络创造条件，为实现全国资源共享作出贡献。

参考文献

1　张玉麟．从手工编目走向计算机编目的新时代．中科院第八次图书馆学情报学科学讨论会文集，1992.

2　刘　荣．图书情报管理自动化基础．武汉大学出版社.