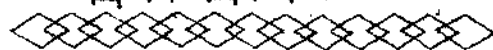


談談机读图书目录 的 編 制

〔刘 彭〕

吉林省图书馆



一九八〇年五月二十九日

谈谈机读图书目录的编制

我国图书馆界近年来相继在研究把电子计算机应用到图书馆工作方面，以期建立自动化的图书情报系统，已见了一些成效。从目前情况来看，用计算机编制西文图书联合目录进而发展为专题的书目文献检索，这可能是我国图书馆应用电子计算机的一个较好的起步点。无论是书目编制系统还是检索服务系统，机读图书目录的编制都是一个重要的环节，也是一个有严格技术规定和比较繁琐的工作环节。

由于研制工作的需要，今年我馆拟定了一个“机读图书目录格式”（初稿），并已按这个格式编了少量机读书目，以便在进一步试验中使用。

谈谈机读图书目录的编制

内容提要

一、对机读图书目录的一些认识	1
1. 机读图书目录与传统手工图书目录的比较	1
2. 对编制机读图书目录的要求	1
3. 对美国国会图书馆机读图书目录《MARCU》的认识	2
二、我馆外文(英文)图书卡片著录情况及其与“马尔克”卡片的比较	4
三、我馆机读图书目录格式的编制	6
1. 说明区	7
2. 目录区	9
3. 资料区	11
4. 我馆机读图书目录可变字段的确定	14
结 语	17

一、对机读目录的一些认识

机读目录就是利用磁、光或电子计算机能阅读的物质上的目录记录。英文简称是Machine Readable Catalog, 简称就是 MARC。“机读图书目录”是“机读目录”中的一种。

1、机读图书目录与传统的手工图书目录的比较：

机读图书目录与传统的^{手工}图书目录(书本式、卡片式)相比较有其相同处和不同处, 它们之间的相同处是:

- ① 工作对象相同 —— 都是书刊资料。
- ② 服务目的相同 —— 是供使用者查找书刊资料的工具。
- ③ 表达的内容相同 —— 也就是它们的著录系统相同, 诸如书刊名称、著者项、出版项、稽核项、附注项、提要项……等, 它们之间的不同处是:

① 载体不同 —— 机读目录的载体是磁带、磁盘等; 传统书目是书本、胶片、卡片等。

② 加工方式不同 —— 机读目录是手工与机器加工相结合传统书目的全部加工过程均是人的手工劳动。

③ 服务方式不同 —— 机读目录必须通过电子计算机来服务; 而传统手工书目是由使用者自行直观运用。

我们还必须认识到, 机读图书目录是在传统的图书卡片式目录基础上为利用电子计算机现代化工具而发展起来的, 但是机读图书目录也自成体系。

2、对编制机读目录的要求

对设计资料档, 机读目录磁带格式在国内和国际都提出了许多要求, 归纳起来有这样几个:

- ① 兼容性: 含意有四个方面, 即是指数据兼容, 格式兼容, 设备兼容和程序系统兼容。数据兼容是指目录格式中要包括

大多数图书馆所需要的数据单元；格式兼容的意思是使设计出的格式能便于进行交换；程序系统兼容，因为磁带是把计算机硬件软件及数据库三者联系在一起的中枢，我们工作目的是使用磁带通过特定的程序来实现的，所以程序系统以磁带格式为手段兼容了这三者相应的要求；改密译码是使目录格式能适应多种计算机和设备的需求。

② 通用性：因为系统不是固定不变的，目录格式就要具有适应系统变化的通用性。而通用性与标准化是分不开的，这方面应有三个标准化内容：标准的磁带目录格式、标准的编码字符集和统一的编目著录规则。

③ 扩展性：机读目录要随着客观条件的发展变化而需不断扩充或改变，由于这种变化要受经济能力、人力及时间性等因素的影响，所以需要逐步的实现，在编制格式时要求尽可能地把手边的扩充内容考虑周到，拟出适当的解决办法。办法一般地有这样几种：

- (i) 确定出备用区（留空白字段）
- (ii) 变动可变字段数
- (iii) 在字段中留用空白字符等。

④ 实用性：编制机读目录（即便是做试验用）一定要考虑它的实用性，要有使用价值，要切合本图书馆的实际情况，要经济节约、由简而繁，从小到大、积少成多。

此外还有一些如经济性、可靠性、保管性等也是需要考虑的因素，然而我们这次是在进行试验研究，所以只把对当前工作影响较大的因素进行了推论。

3. 对美国国会图书馆机读图书目录《MARC II》认识

《MARC II》系统是美国国会图书馆继《MARC I》之后于一九六八年六月完成的。由于它经过较长时间的试验和修正充

从而具有它自己的特長和長处，才使它一經問世就显出了强大的生命力。許多拉丁語系和非拉丁語系的國家的圖書館都先後接受MARC系統，並在其基礎上發展為本國的機讀目錄系統，如英國的《UKMARC》、日本的《JMARC》和國際機讀目錄格式《UNIMARC》（正在討論修改中），我國在這方面也正在進行工作。

“馬爾克”（MARC的中譯音）的長处和特長是什麼呢？

首先，“馬爾克Ⅰ”機讀目錄是經過大約3年時間的實踐和不斷改進，經歷了《MARCI》和《MARCIⅡ》兩個階段才最後臻於完成，而在以後的使用過程中仍在不斷地總結提高。

其次，在編目方面採用的是國際上通用的標準著錄規則。它的目錄體系採納了《英美編目系列》（AACR），這一條例是國際編目規則會議通過且已被多數國家所採用的，它在描述性著錄格式方面採用了《國際標準著錄格式（ISBD）》。

第三，“馬爾克Ⅱ”磁帶採用的是美國國家標準學會通過的磁帶位應交換標準字符集，即美國位應交換標準代碼（ASCII）——此碼是國際標準化組織（ISO）所承認的。並將這個標準的代碼由原來的7位代碼組成，扩充到國際上通用的九磁道磁帶用的8位代碼和七磁道磁帶用的6位代碼。

“馬爾克”目錄發行的是九磁道和七磁道半吋磁帶。使用的是COBOL和FORTRAN國際通用語言。所以它對不同型號電子計算機有較強的適應性（但我國國內則是通用1吋磁帶）。

第四，“馬爾克Ⅱ”在格式中保留了圖書館的傳統著錄內涵和卡片目錄體系。它能把傳統的卡片目錄內容全部記錄到磁帶上，輸入到機器中，代替手工機序的卡片目錄。它既能發揮電子計算機的大量儲存圖書目錄及快速檢索的優點

性，又保持了图书馆的传统目录体系。

第五，它的系统在使用中可以根据各图书馆目录体系的需要进行修改扩展和补充，这就是它把保持本身的固有体系和适应各个方面的灵活性较好的统一起来了。

二、我馆外文（英文）图书卡片著录情况 况及与“马尔克”卡片的比较

机读图书目录是在现用的传统卡片基础上编制的，故在确定机读图书目录格式时首先要对现用的图书卡片著录项目进行分析。因我们目前研试以英文书目为对象故本文分析亦以英文图书卡片为例。

我馆现用的英文卡片是采用北京图书馆统一编制的铅印卡片，部分无统编的卡片者由馆自己编制打印，但著录项目风格式也与统编的相同。卡片上著录项目一般有以下各项：

- ① 著者项（基本著录）
- ② 书各款项（包括题下项——付书名）
- ③ 出版项
- ④ 稽核项
- ⑤ 附注项或说明项：卡片上关于图书分类有按《中国图书馆图书分类法》分类的分类号 and 按《中国科学院图书馆图书分类法》分类的分类号；还有按《卡特分类号码表》（C.A. CUTLER《ALPHABETIC-ORDER TABLE》）分编的著者号；此外还有书价（人民币）、征订目录号等。使用这个卡片时补编入我馆的分类号及著书号即可付用。我馆是按《中国图书馆图书分类法》分类，著书号分为四部分：①文种号 ②类号 ③书次号 ④卷次号

（卡片样式见图一）

与“马尔克”卡片格式相对照，不难看出它们之间是基本上一致，略有差别（“马尔克”卡片样式见图二）现比较如下：

① 它们的基本著录一致，都是取著者标目为基本著录。当然在具体技术处理上不尽相同。

② 名著录项目的排列和次序也极其相似，如著者项，书名项、出版项、稽核项、附注项等。

③ “马尔克”系统除主要著录项外，尚有丛出项（有作根查用和不作根查用的）、附加项（主要是补充主要款目中之不足，比如主属附加项目、丛书附加项目，……等等）我们现用的卡片目前还有这些项目。

④ 使用的分类法不同，“马尔克”使用的是“杜威分类法”。

由于著录项目和内篇方面同多异少，从而使我们参照“马尔克”格式制定我们的格式有了方便条件。

★ 41	Kaufmann, Arnold et al.
TP302	Mathematical models for the study of the reliability of systems. New York, Academic Press, 1977
8	221p. illus. Ref
	(Mathematics in science and engineering, 124)
TP 302	
	T3.6722/K45
	F108/23 (1.30)

图一、我馆用的图书卡片格式

Sugarman, Stephen		
Petroleum industry handbook (Edited by Stephen Sugarman, n.p.)		
Published by J.M. Weiner for D.H. Blair (1969)		
xxii, 794p. illus., maps. 29cm.		
"For limited distribution only."		
1. Petroleum industry and trade--Handbooks, manuals, etc. I.		
Title. HD 9560.5.58	338.2'T282	75-1048
MARC		

图二 "MARC" 图书卡片格式

三、我馆机读图书目录格式的编制

目前我国标准的交换磁带记录格式尚未公布，这给我们的磁带格式增加了困难。但是我国已有许多单位都在进行这方面的工作，有的是把国外磁带（包括《MARC II》）引进直接利用，有的是在参照引进的磁带（主要是《MARC II》）格式研制自己的磁带格式。我们在分析了有关资料和学习兄弟单位经验之后，认为可以基本采用《MARC II》图书磁带格式，参照国际标准化组织的《ISO-2709》（1973）和《DP-2709-REV1》（1978）以及若干资料，结合我馆目录情况及满足试验任务的需要来制定我们自己的试验格式。我们给它定名为《吉林省图书馆机读图书目录格式》（初稿）（以下简称“格式初稿”）。

格式初稿结构中，把纵的层次定为二级：区——字段，根据需要某些区下还派生有“分区”，字段下派生有“子字段”。

即区——(分区)——字段——(分字段)。

格式初稿中把一个资料纪录(对应的即是一张图书卡片)分为说明区、目录区、资料区等三个区(图三)

说 明 区	目 录 区	资 料 区
-------	-------	-------

图三 机读图书目录格式结构

1. 说明区:

说明区也称记录头标、领词等(leader)本区是用于存放资料纪录中的数据性质、目录部分的描述形式等的综合性数据它对于一个资料纪录起开头及性质、内容形式说明的作用。是由24个字符组成的固定长字段,总共还有八部分,每个部分均有一定的比例、一定的字符数和固定的字符位置。

0-4 5 6 7 8-9 10 11 12-16 17 18-23

配 录 总 长	配 录 状 态	资 料 类 型	资 料 级 别	空 白	各段符 号的数 量指目 录系	分 段 化	码 数 目	数 据 基 址	资 料 编 址	完 查 情 况	备 注
------------------	------------------	------------------	------------------	--------	-------------------------	-------------	-------------	------------------	------------------	------------------	--------

资料性质

图4 说明区结构

① 记录总长:(第0—4字节)是说明区的首部,也是整个资料记录的首部,它以数字表示从资料记录第一个字符起到结束的“资料记录结束符”止的全部字符数,共5位数字占用5个字节,整个设计的格式允许每个资料记录最多可容纳10万个字符。

~ 8 ~

②. 记录状态: (第5个字节)这是说明书目著录的状态的,如是新的或是修改过的(新书采用字母“N”表示,修正过的用“C”表示)佔1个字节。

③. 资料性质: (第6—9字节)是说明本书目资料若干方面的性质和类型的,需用二个单项说明,一个是资料类型(佔第6字节),例如用“a”代表印刷的文学资料如图书,用“b”代表手写的文学资料如手稿等等,另一个是资料级别(佔第7字节)例如用“m”表示专著,用“s”表示文集等等。第8—9字节暂时空白待用。

④. 各可变字段指示符数目: (第10字节),在机读目录中每个可变字段中,在资料数据前面的字段开头两个字符规定为指示符(每个可变字段里的指示符都有各自的含义和作用)因而指示符数目代码为“2”。

⑤. 分字段代码数目: (第11字节)在可变字段中的每个资料数据单元(或一个独立的词意单元)都组成一个分字段,每个分字段之前也需要用二个字符组成的分字段代码来标识(每个分字段代码也有它自己的字符和含义)故分字段代码也为“2”。

⑥. 数据基地址: (第12—16字节),数据基地址是资料区里第一个字符的位置,它是资料区中各个字段编制“起始位置”数字码的起点,每个资料字段起始字符数要从基地址为要起算的。基地址数等于说明区和目录区的长度总和。

⑦. 资料编码完整情况: (第17字节)本项表示资料记录的完整程度。例如当编制的书目著录项目是查阅书本上实际资料而编写的则以“b”(空白符)表示,称为“完全级”,如不是查阅实际书本资料而编写的称为“不完全级”以字母“L”表示。

③ 备用空位：(第18~23字节)这些字节在《MARC II》及“ISO2709”中有的已被使用，但考虑我国实际情况暂时不用，待扩充时备用。

在编制机读图书目录时，说明区里的各个项目，要在目录区和资料区编制完成之后才能进行编制。

2. 目录区：

又称“记录的地址目录”，它是由表示资料区中各资料数据字段的种类、位置、长度三个项目的数字码组成的区域。它是一个机读目录内容的目录表（与一本书的目录表相似）。在资料区中有多少个“资料字段”，在目录区中就有多少个“目录字段”与其按顺序相对应，每个“目录字段”代表一个“资料字段”。这种把“目录字段”集中起来的放在资料区之前形成目录区的表示方法在机读目录编制中被称为“标识——值”编组方式。

“目录字段”是由三个数字码组成的固定长字段，但目录区是可变长字段（因为资料区内的可变字段数目不定）。目录字段下分标识，字段长度和起始字符位置三部分。

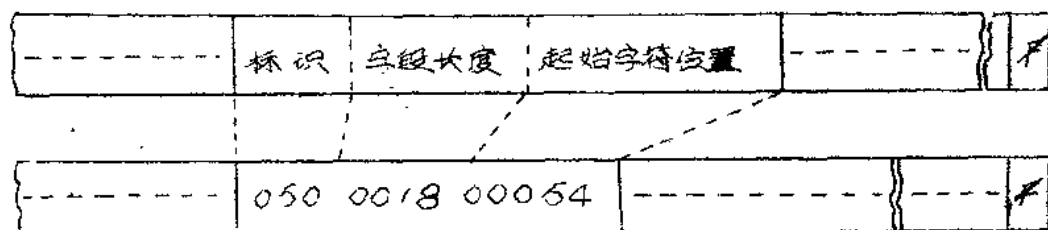
① 标识：主要是指可变字段的标识，是表示资料区存放的资料字段的类目的，它由三个数字码组成，在“马尔克”中或各种磁带格式中都有个自的规定，我们是参照“马尔克”的标识表制定的标识，在一种格式中标识确定之后要正确使用，不能任意变更。标识在资料目录字段中佔头三位数字。

② 字段长度：表示所标识的资料字段的总的字符数。注意：是指该资料字段中的全部字符数，包括指示符、分字段代码、数据本体、分隔符、字段结束符等的总和数。字段长度在资料目录字段中继标识的三位数字之后佔四位数字。

③ 起始字符位置。这是指所标识的资料字段的第一个字符的位置。这个位置数字是从数据基地址起标的字符位置数（而不是从格式记录的头标的第一个字符起算）。

应该指出的是，在资料区中各可（变）字段（资料字段）中实际资料数据起点是该字段中的第五个字符，头四个字符里的前两个是指示符后两个是分段代码。

起始字符位置在目录字段中佔最后五位数字。



图五、目录字段结构及编码实例

（F：字段结束符）

无疑在编制目录时，要待资料区完成才能进行。目录区完了，在最后加一个字段结束符“F”。

3. 资料区:

资料区是〈格式〉初稿里的主体区，卡片式目录的着录项目都通过一定的固定格式转换到资料区中贮存。前述区的说明区及目录区在格式中都是为适应电子计算机的功能需要而为资料区服务的，也是为系统内部服务的，对外界起直接服务作用的是资料区。

为方便起见，把资料区又分为二个分区：“控制字段分区”和“可变字段分区”。

控制字段分区:

按我们当前具体情况，控制字段只用标识号为“006”的“固定长的数据单元”一种，它是一个具有40个字符的固定长字段。它通过一些具有固定含义的代码和数字，把资料区内各资料字段的特征和概要状况给指示出来。我们拟定格式初稿时，系参照“马尔克”图书目录中相应的字段，采用了其不变改结构的一部分款目，对其中某些当前尚没必要用的款目暂时不用，其字符位置待用。格式初稿中启用十三个部分，使用34个字节，空白6个字节。它的具体组成安排如下：

数据单元名称	字符数目	在字段中的字符位置
入档日期	6	0 ~ 5
出版日期的类型代码	1	6
出版年1	4	7 ~ 10
出版年2	4	11 ~ 14
出版国家代码	3	15 ~ 17
刊号代码	4	18 ~ 21
空白	1	22
编制形成代码	1	23
内容形式代码	4	24 ~ 27

~12~

空 白

28

会议出版物指示符

1

29

纪念文集指示符

1

30

空 白

2

31 ~ 32

小说指示符

1

33

传记指示符

1

34

文种代码

3

35 ~ 37

空 白

2

38 ~ 39

有款目内各代码及其用法可参见格式初稿，本文不译述。

字段

位置 0~5 6 7~10 11~14 15~17 18~21 22 23 24~27 28 29 30 31 32 33 34 35~37 38~39

八	出	出	出	出	插	空	复	内	空	会	纪	空	小	传	文	空
相	版	版	版	版	图		制	各		兴	念		说	记	种	
目	日	期	年	年	代	白	形	形	白	出	文		指	指	代	白
期	期	期	年	年	代	白	式	式	白	版	集	指	示	示	码	白
	代	代	代	代	代	代	代	代	代	物	指	示	符	符		
	码	码	码	码	码	码	码	码	码	指	示	符				

图六、“008”控制字段结构

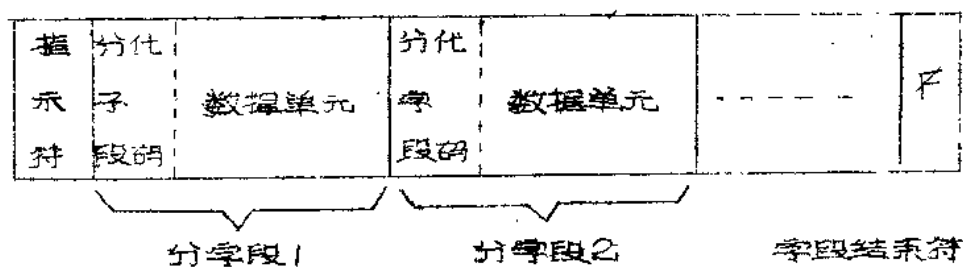
在字段末尾加一个字段结束符“#”

可变字段分区：

可变字段就是“资料字段”，因为图书卡片上的各著录项的内名词汇数量不一，所以采用了可变长字段，这样可比较灵活地适应著录项的需要。

整个可变字段分区内是用若干组独立的可变字段组成；每组字段都相应的表示图书卡片上一个著录项目的内容。由于每张图书卡片上除了一些少量的著录项目外，尚有一些根据个自的需要编入的项目，所以可变字段区内的可变字段组总数不是固定不变的。

每组可变字段都由指示符开始，在指示符之后就是组成本字段的一组或若干组分字段，其数量由具体项目的内容决定。每组可变字段的末端都有一个字段结束符“F”。



图七 可变字段结构

最后一组可变字段完了，若这个字段后面再没有其它数据段，这就 明整个资料记录结束了，这时，在最后一个字段结束处不加字段结束符“F”而加一个记录结束符“R”。

(1)指示符：由两个字符组成，在某种情况下还可以使用空白符“ ”， 指示符的规定可见格式初稿中的附表。

(2)分字段代码：由两个字符组成，第一个字符是分隔符“\$”，第二个字符是小写的英文字母，它既起分字段间分隔作用，又起分字段标识作用，详见格式初稿说明。

(3)分字段数据单元：由不定长数据组成，它取决于图书卡片上的对应项目的组成情况。数据或字符紧接在分字段代码之后，数据中各个独立词之间要用一个空白符“ ”分隔开。

(4)字段结束符：每组可变字段完了都要用一个字段结束符

“#”，它也起着可变字段间的间隔作用。

(5)记录结束符：在整个目录记录完成后，于最后一组可变字段之末尾記入一个记录结束符“R”，表示本记录结束。

(6)关于可变字段的“标识”：每个可变字段都有代表该字段的通用标识，用以记在目录区里的“目录字段”中，标识号在本可变字段中並不出现。

在“马尔克”里有可变字段标识表，里面编录了各种项目的标识号。根据我馆情况，在编制格式初稿时选用了一部分，並也补充了一部分。标识号是由三个阿拉伯数码组成。

4. 我馆机读图书目录可变字段的确定：

当卡片式目录上的著录项目固定下来，机读图书目录格式的可变字段也就随之基本上确定了。我馆格式初稿中，可变字段确定为七种十个标识号的资料字段，见下表：

标识	字段内容
(1) 090	吉林省图书馆藏书分类号及片号
(2) 主要款目	—— 著者项
100	个人姓名
110	团体名称
111	会议或集会
(3) 245	书名项
(4) 250	版次项
(5) 260	出版项
(6) 300	稽核项
(7) 附注项	
500	一般性说明
505	内容说明

注：详细情况请参见格式初稿