

中文國際化問題與國際化標準問題 Globalization Issues Related For Chinese Language

中文國際化問題與國際化標準問題 Globalization Issues Related For Chinese Language

國際化標準問題

國際化標準問題

中文拓片机读目录格式使用手册

MARC Format Manual For Chinese Rubbings

中文拓片编目规则

Cataloging Rules For Chinese Rubbings

中国国家图书馆 编

北京图书馆出版社

图书在版编目(CIP)数据

中文拓片机读目录格式使用手册·中文拓片编目规则/中国国家图书馆编. —北京:北京图书馆出版社,2002.1

ISBN 7-5013-1877-8

I. 中… II. 中… III. ①拓片—机器可读目录—中国—目录使用—手册②拓片—编目规则—中国 IV. ①G256.364②G254

中国版本图书馆 CIP 数据核字(2001)第 087883 号

书名 中文拓片机读目录格式使用手册·中文拓片编目规则

著者 中国国家图书馆 编

出版 北京图书馆出版社 (100034 北京西城区文津街7号)

发行 (010)66126153 传真 (010)66174391

E-mail Btsfxb@publicf.nlc.gov.cn

经销 新华书店

印刷 北京飞达印刷厂

开本 787×1092 毫米 1/16

印张 7.25

版次 2002 年 1 月第 1 版 2002 年 1 月第 1 次印刷

字数 130(千字)

印数 1—3000

书号 ISBN 7-5013-1877-8/G·502

定价 35.00 元

前 言

书目记录的规范化、标准化是实现文献资源共享的前提。在这方面,中文普通图书和古籍早已进行了尝试并取得了可喜的成果,中文拓片却起步较晚。

传拓技术的出现可以追溯到南北朝时期或者更早的汉代,拓片的最早记录是《隋书·经籍志》,北宋以来,碑帖书目大量出现,然而书目的著录却很不规范。至今,中文拓片的著录还没有一个国家标准。1999年,我们结合中文拓片的特点,依据《中国机读目录格式使用手册》(北京图书馆《中国机读目录格式使用手册》编委会编,华艺出版社1995年12月出版),参照中文普通图书机读目录格式和中文古籍机读目录格式、《中国文献编目规则·金石拓片部分》(黄俊贵主编,广东人民出版社1996年出版)编写出了《中文拓片机读目录格式使用手册》初稿,并在实际工作中予以检验;2000年上半年对其进行修改,11月召开了国内有关专家和同行的座谈会征求意见;根据专家的意见和修订后的《中国机读目录格式使用手册》(潘太明等修订,科学技术文献出版社2001年3月出版)再次进行修改,并于2001年10月征求有关专家的意见后修改完稿。

中文拓片与其他中文书籍有很大的差异,因此,我们本着求大同存小异的原则对一些字段进行了变通。本手册是针对中文拓片而制定的,对影印件视情况不同可依中文图书和中文古籍机读目录格式编定。

为了便于手册的使用,我们将修订后的《中文拓片编目规则》(本规则原收辑于《中国文献编目规则》)一并出版。

手册和规则的编写、修订始终是在馆业务处和善本部领导的指导下进行的,并得到了北京大学、中国科学院图书馆、中山大学图书馆等国内同行的大力支持。在编写、修订过程中,承蒙周升恒、朱岩、汪东波、富平、曹宁、延卫平、苑克丽、姚伯岳、罗琳、胡海帆、李庆涛、申军等专家的热情帮助,在此,谨向他们深表谢意。参加此项工作的还有金石组的贾双喜、吴元真、王丽燕、袁玉红同志。

由于我们的工作时间短,经验有限,手册与规则中还有很多问题需要在实际工作中进一步验证和完善,敬请业内外人士批评指正。

编者

2001年12月

目 录

前言	(1)
----------	-----

中文拓片机读目录格式使用手册

一、总论	(3)
二、记录头标和数据字段——详细说明	(6)
记录头标	(6)
0-- 标识块	(11)
001 记录标识号	(12)
005 记录处理时间标识	(13)
010 国际标准书号 (ISBN)	(13)
035 源记录标识号	(14)
1-- 编码信息块	(15)
100 通用处理数据	(16)
101 作品语种	(18)
191 编码数据字段:拓片	(19)
195 编码数据字段:刻立时间	(23)
2-- 著录信息块	(25)
200 题名与责任说明	(25)
205 拓片版本说明	(28)
210 出版发行	(28)
215 载体形态	(29)
225 丛编	(30)
3-- 附注块	(31)
300 一般性附注	(32)
302 编码信息附注	(33)
304 题名与责任说明附注	(33)
305 拓片版本附注	(34)
306 原石特征等附注	(35)
307 拓片形态特征附注	(36)
310 装订及获得方式附注	(36)
311 连接字段附注	(37)

312	相关题名附注	(38)
314	知识责任附注	(39)
316	在编复本的附注	(39)
317	出处附注	(40)
318	保护操作附注	(40)
320	书目、索引附注	(41)
321	被索引、摘要和引用附注	(41)
327	内容附注	(42)
330	全文	(43)
345	采访信息附注	(43)
393	系统外字附注	(44)
4--	款目连接块	(44)
410	丛编	(46)
422	正编	(46)
461	总集	(47)
462	分集	(47)
463	单册	(47)
482	和...一起装订	(48)
5--	相关题名块	(48)
500	统一题名	(49)
510	并列正题名	(50)
512	封面题名	(51)
514	卷端题名	(52)
517	其他题名	(53)
540	编目员补充的附加题名	(55)
6--	主题分析块	(56)
610	非控主题词	(56)
620	出版地/制作地检索点	(57)
7--	知识责任块	(58)
701	个人名称——第一知识责任	(58)
702	个人名称——其他知识责任	(60)
711	团体名称——第一知识责任	(61)
712	团体名称——其他知识责任	(63)
8--	国际使用块	(64)
801	记录来源	(64)
856	电子文件地址与检索	(65)
9--	国内使用块	(69)
905	馆藏信息	(69)

中文拓片编目规则

0 通则	(73)
1 题名与责任说明项	(76)
2 版本项	(82)
3 文献特殊细节项	(82)
4 刻立项	(82)
5 载体形态项	(84)
6 丛编项	(85)
7 附注项	(86)
8 标准编号、装裱与获得方式项	(87)
附录:	(89)
一、中文拓片关键词表	(89)
二、世界语种代码	(91)
三、中文拓片机读目录格式字段表	(102)

中文拓片机读目录格式使用手册

一、总论

1. 编制目的

中文拓片机读目录格式为中国各书目机构与其他国家书目机构之间以及中国国内拓片典藏和使用单位之间交换中文拓片书目数据提供标准的计算机识别格式。

该格式在中文拓片书目数据规范方面为中文拓片书目数据库的建立和中文拓片书目数据处理提供参照或依据。

本使用手册对中文拓片机读目录格式及其实际应用作了明确说明,并提供各类典型示例,使根据本格式编制的中文拓片书目数据标准化、规范化,符合 MARC 格式要求,能够进行国际国内的数据交流和资源共享。

2. 文献涵盖范围

该格式涵盖的文献是中文(含中、外文合璧)拓片。

3. 依据标准

该格式的主要依据:

UNIMARC Manual/IFLA2000 年修订版。

WH/T0503—96 中国机读目录格式/中华人民共和国文化部提出。—北京:中华人民共和国文化部,1996.7.1 实施。

4. 拓片机读目录格式的总体规定

本格式记录遵照 ISO 2709(GB 2901)标准。每个记录由记录头标区、地址目次区、数据字段区和一个记录分隔符组成。

记录结构

记录头标	地址目次区	数据字段区	记录分隔符
------	-------	-------	-------

记录头标:24 个字符长,由定长数据元素构成,包含记录处理所需数据。

地址目次区:由若干个目次项和一个字段分隔符组成。每个目次项由三个数字的字段号、四个数字的字段长度和五个数字的起始字符位置构成。

目次项 1

目次项 2

字段号	字段长度	起始字符位置	……	……	字段分隔符
-----	------	--------	----	----	-------

数据字段区:由若干固定长和可变长字段构成。除 001 和 005 字段无指示符和子字段标识符外,其余字段都具有如下结构:

指示符

第一个子字段

其他子字段

指示符 1	指示符 2	\$*	数据	……	字段分隔符
-------	-------	-----	----	----	-------

拓片机读记录标识符号与 UNIMARC Manual/IFLA2000 年修订版一致。

“#”表示空格。

5. 机读记录的功能块

机读记录分为若干个功能块。字段标识符的第一个数字(最左位)表示字段所属的功能块。

- 0-- 标识块:记录一些标识该记录的文献标识号。
- 1-- 编码信息块:包含一些描述记录或数据各个方面的固定长编码数据。
- 2-- 著录信息块:记录 ISBD 所规定的著录项目(附注项除外)。
- 3-- 附注块:包含一些自由行文的附注项,以补充说明著录、检索点以及其他内容。
- 4-- 连接款目块:包含一些揭示该记录与其他书目记录之间的关系的字段。这些关系包括:层次关系、平行关系和时间关系等。
- 5-- 相关题名块:包含一些编目拓片正题名以外的并通常出现在拓片上的其他题名。
- 6-- 主题分析块:包含由词语构成的不同体系的主题数据。
- 7-- 知识责任块:记录对所编拓片负有某种知识责任的个人和团体名称。
- 8-- 国际使用块:包含一些不宜记入前面各块(0--至 7--)由国际统一使用的字段。
- 9-- 国内使用块:这块定义字段为中国使用。因此,国家格式可以自行定义的字段范围有:9--字段、-9-字段、-9 字段,以及子字段 \$9 和指示符 9。

6. 机读记录的必备字段

中文拓片机读记录除记录头标区、地址目次区外,还必须提供以下必备字段:

- 001 记录标识号
- 100 通用处理数据
- 101 拓片语种
- 191 编码数据字段:拓片
- 200 题名与责任者项(其中 \$a 正题名为必备子字段)
- 7-- 知识责任块
- 801 记录来源
- 856 电子文件地址与检索

7. 不同文字的处理

在字段 001—099 和 200—899 中,使用适当的内容标识符来记录可替换的图形文字,如用汉语拼音表示题名。为了记录可替换图形文字(无论是否出现在作品中),格式中的所有字段均认为是可重复的。那些被列为不可重复的字段,对于记录中的每一种可替换的图形代码或文字来说只能重复一次。

该技术为编目机构按标准规则记录的拉丁化文字、音译文字和变体文字或表音文字提供了一种途径。

在每个记有替换图形代码或文字的重复的字段中,包含子字段 \$6(字段间连接数据),如果需用的话,还含有子字段 \$7(字段的字母/文字)。 $\$6$ 和 $\$7$ 子字段的具体使用说明如下:

\$6 字段间连接数据

本子字段含有该字段向记录中其他字段连接的数据,还包含一个说明连接原因的代码。当使用本子字段时,本子字段的前两个元素(字符位 0—2)必备;第三个元素(字符位 3—5)是可选择的。所以本子字段的长度可能是 3 个字符,也可能是 6 个字符。子字段 \$6 通常定为字段的第一个子字段,在 \$7 前面,不可重复。子字段 \$6 的数据如下:

数据元素名称	字符数	字符位置
(1)连接说明代码	1	0
(2)连接号	2	1—2
(3)被连接字段的标识符	3	3—5

(1)连接说明代码(字符位置 0)

这个代码标示字段间连接的原因。

a = 替换图形表示或文字(如汉语拼音)

z = 其他

(2)连接号(字符位置 1—2)

为两个数字号码,记入要连接在一起的两个字段的 \$6 子字段里。连接号的作用是为了匹配相关的字段(它不起顺序号或地址号的作用)。连接号可以随意分配,只要连接号码在每对连接字段内是一致的(并且是唯一的),就可以随机赋值。

(3)被连接字段的标识符(字段位置 3—5)

这个数据元素为被连接字段的字段标识号。可选择使用。如果相互连接的两个字段的标识号相同,则通常省略这个元素。

\$6 子字段可在 5--(500 字段除外)、7--等可重复字段中使用。

\$7 字段的字母/文字

本子字段含有表示字段主要内容的字母或文字的代码。代码值使用 100 字段字符位置 34—35 题名字母所定义的代码。当本字段的代码与 100 字段 34—35 字符位所取的字母或文字的代码相同时,本子字段通常是省略的。

本子字段应直接置于字段的第一个数据子字段(如 \$a)的前面。通常在 \$6 之后,除非不存在并列字段(\$6 子字段不出现)。

按照 UNIMARC 格式的规定,汉语拼音属于交替文字形式,应采用重复字段的方法,用 \$6 子字段来连接,用 \$7 子字段指明其文字形式(即 ba = 拉丁文字)。在我国,中文图书机读书目已习惯使用 \$A 或 \$9 来标识汉语拼音,但由于这种方法不能用于数据的国际间交换,本格式建议使用重复字段的方法来处理汉语拼音。

对个人名称、团体名称、题名等,应给出单个汉字的汉语拼音。单个汉字拼音之间以一空格分隔,并一律采用小写、单字节表示。汉语拼音中不使用任何标点符号。在汉语拼音字段,数字、外文字母照录,外文字母一律采用大写,外文字母和数字里的标点符号照录不省略。

8. 内容结构

本手册的正文为中文拓片机读目录格式的使用手册。

本手册的附录包括:1. 中文拓片关键词表;2. 世界语种代码;3. 中文拓片机读目录格式字段表。

二、记录头标和数据字段——详细说明

记录头标

1. 定义

本部分包含根据 ISO 2709 制订的对中文拓片书目记录进行处理时所需的通用信息。

2. 出现情况

记录头标出现在每个记录开头,它是必备的和不可重复的。

3. 字段号、指示符和子字段

记录头标没有字段号、指示符和子字段标识符。

4. 定长数据元素

头标中的数据元素是由字符位置标识的。头标的总长度为 24 个字符。字符位置规定从 0 至 23。

定长数据元素表

数据元素名称	字符数	字符位置
(1)记录长度	5	0—4
(2)记录状态	1	5
(3)执行代码	4	6—9
(4)指示符长度	1	10
(5)子字段标识符长度	1	11
(6)数据起始地址	5	12—16
(7)记录附加定义	3	17—19
(8)地址目次区款目结构	4	20—23

5. 内容注释

记录头标定位在每个记录的开头,它包含处理记录所需的数据。字符位 9,10,11,20—23 所含的特定值可通过计算机程序生成。字符位 0—4 和 12—16 所含的是数字型数据,它表示记录中相应部分的字符数量可能由计算机在生成记录时计算产生。字符位 5,6—9,17—19 的值,如果采用外来源数据,可以从源记录数据中通过转换程序生成。若使用本格式制作源记录,则由手工指定。

(1) 记录长度(字符位置 0—4)

五个十进制数字,右对齐,空位必须填0(零)。表示整个记录的字符总数,包括记录头标、地址目次区和数据字段区三大部分和记录分隔符。当为了交换而组装成 MARC 记录时,该项数据通常由计算机自动生成。

(2) 记录状态(字符位置 5)

一位代码表示记录处理状态。

c = 经修改的记录

当由于改错而变更成为一个新记录,或者记录中有字段被删除,均用“c”。若整个记录由后来发行的记录代替,也要用“c”。因此,一个已标定了“n”、“o”的记录若做了更改,代码均应为“c”。

d = 被删除的记录

为了标识一个记录不再有效,可以通过提供该记录控制号的方法进行交换。这类记录可以仅包含记录头标、地址目次区和 001(记录控制号)字段,或者包含记录发行时的全部字段。无论哪种情况,都可使用 300 通用附注字段说明该记录被删除的原因。

n = 新记录

新记录均用“n”标识。如果代码“o”适用,则不用“n”。

o = 较高层新记录

在分层体系中,低于最高层的较高层新记录用“o”表示(参见第 8 字符位即层次等级取 2)。

通常一个拓片记录的记录状态在“新记录”、“较高层新记录”、“经修改的记录”和“被删除的记录”当中选一个。准备正式写入数据库的新记录选“n”、“o”。已正式写入数据库发行后又修改或删除的记录则选“c”或“d”。

(3) 执行代码(字符位置 6—9)

本格式采用的 4 位代码定义如下:

(3a) 记录类型(字符位置 6)

拓片记录常用的记录类型代码如下:

a = 文字资料印刷品(拓片)

b = 文字资料手稿(墨迹)

拓片属于印刷品的范畴,记录类型取代码“a”。

例 董美人墓志/(隋)杨秀撰

本例拓片之外有张廷济等六人墨跋。拓片的“记录类型”选“a”,墨跋子目的“记录类型”选“b”。

(3b) 书目级别(字符位置 7)

拓片记录常用的书目级别代码如下:

m = 专著

专著是指单题名的单幅、卷轴装拓片或丛刻类拓片实体。

例 1 董美人墓志/(隋)杨秀撰

单幅拓片书目级别为“m”。

例 2 东书堂法帖:10 卷/(明)朱有燬摹

丛帖属丛刻类拓片,书目级别为“m”。

a = 分析级

该书目所描述的实体在物理上包含在另一实体之中。如丛帖子目、拓片子目(拓片的重要序、跋、题记等需要单独编子目)等。

例1 东都帖/(唐)李世民撰并行书

此为《东书堂法帖》的子目之一,书目级别为“a”。

例2 董美人墓志跋/(清)张廷济撰

此为张廷济为《董美人墓志》作的跋文,单编子目其书目级别为“a”。

c = 合集

汇集型拓片实体。即不同版本的拓片合装成册,冠有或不冠有统一题名的集拓本、合裱本;多个不同题名的拓片合拓或合裱一纸的单幅拓片。

例1 陈介祺藏金文拓本/(清)陈介祺藏

此为各种不同时代、类型的拓片合裱成20册,其书目级别为“c”。

例2 妙义造像记·侯三娘造像记/[佚名]撰

此为龙门石窟造像记,两造像记合拓一纸,其书目级别为“c”。

记录的书目级别与记录的主体部分有关,或者说与该记录描述的主要书目实体有关,该实体的题名出现在200字段里。

(3c)层次等级代码(字符位置 8)

该代码标示本记录以层次性的关系与其他记录连接,并表明它在层次中的相对位置。该代码旨在为系统使用,以便在层次结构中形成各个记录。层次结构对被连接的各记录并不提供足够的信息,因为被连接的其他记录总是存贮在同一个数据库中。所用代码如下:

= 层次关系未定

0 = 无层次关系

指不含有分析记录具有独立存在形式的拓片实体。

1 = 最高层记录

指含有分析记录且具有独立存在形式的拓片实体。如本身有分析记录的单幅拓片、丛拓、集拓等。

2 = 最高层以下层次的记录(所有低层记录)

指所有分析记录。

本手册要求拓片记录只作该拓片向其上一层拓片的连接,而不作高层向低层的向下连接,因此,只要记录中出现4--字段,此代码必须选“2”。

有的机构不做记录间涉及层次的连接,则该代码总取“#”。如果字符位5为0,则字符位8应取2。代码“1”和“2”仅用在其他级别的被连接记录事实上已经存在的情况下。

(3d)未定义(字符位置 9)

含空格“#”。

(4)指示符长度(字符位置 10)

一位十进制数,标示字段指示符长度。本格式取2。

(5)子字段标识符长度(字符位置 11)

一位十进制数,标示子字段标识符(如\$a)的长度,本格式取2。

(6)数据基地址(字符位置 12—16)

五位十进制数,右对齐,不足五位时空位补零(0)。该地址标示第一个数据字段相对于记录首位的起始字符位置。由于记录的第一个字符定义为“0”位。所以记录的数据基地址数等于头标和地址目次区(包括该区末尾的字段分隔符)字符数量的总和。在地址目次区中,每个字段的起始字符位置是相对于第一个数据字段的第一个字符的位置,并不是从记录的首位算起。每个字段位置都可根据基地址算出。该基地址数在组装 MARC 记录时自动生成。

(7)记录附加定义(字符位置 17—19)

三个字符位所含代码给出了处理记录所需的细节。

(7a)编目等级(字符位置 17)

一位字符代码,通常标明机读记录的完整程度以及编制记录时是否与原编目实体核对过。

= 完全级。编制机读记录时,与原编目实体核对过。

1 = 次级 1。编制机读记录时,未与原编目实体核对过。如:该记录是根据目录卡片制作的,其字段号、指示符和子字段代码可能不具有与原编目实体核对过的那种可靠性和准确性。

(7b)著录格式(字符位置 18)

用一位字符代码标出该记录采用的著录规则。说明在编制 200—225 字段时是否依据了国际标准书目著录(ISBD)的规定。其代码如下:

= 记录完全采用 ISBD 格式:记录中出现的全部 ISBD 数据元素都符合 ISBD 规定。

i = 记录为部分的或不完全的 ISBD 格式:不是全部而是部分字段符合 ISBD 规定。

n = 记录为非 ISBD 类格式:记录中的数据元素不符合 ISBD 规定。

(7c)未定义(字符位置 19)

含一位空格“#”。

(8)地址目次区款目结构(字符位置 20—23)

此区为每一个字段的目次款目的长度和结构提供了详细的说明。四个字符位的内容定义如下:

(8a)“字段长度”的长度(字符位置 20)

一位十进制数,标示每个地址目次款目里“字段长度”部分所含字符数。本格式取值为 4。

(8b)“起始字符位置”长度(字符位置 21)

一位十进制数,标示每个地址目次款目里“起始字符位置”部分所含字符数。本格式取值为 5。

(8c)“执行定义部分”长度(字符位置 22)

一位十进制数,标示每个地址目次款目里“执行定义部分”所占字符位数。由于本格式地址目次区中不含这部分,故本格式取值为 0。

(9)未定义(字符位置 23)

填空格“#”。

6. 相关字段

记录头标中提供的数据元素在格式的其他部分是没有的。虽然执行代码的一些值,如“记录类型”和“书目级别”与其他代码数据有些相重,但事实上头标里的代码表示的是记录的

特征,而不是直接表示编目实体本身的特征。

7. 实例

例 1 00816nam0#2200265###450#

本例表示的是一个单张拓片新记录的头标。该记录的制作符合 ISBD 的规定,并与原拓片核对过。其构成如下:

字符位	值	注释
0—4	00816	整个记录的字符总数为 816 个,前两位字符填 0。
5	n	新记录。
6	a	记录类型—拓片。
7	m	书目级别—专著。
8	0	层次等级—无层次关系。
9	#	未定义,填空格。
10	2	指示符长度—总为 2。
11	2	子字段标识符长度—总为 2。
12—16	00265	记录头标和地址目次区的字符总数为 265 个,即数据基地址(数据紧接在目次区之后)为 265。
17	#	编目等级—完全级:制作记录时,与原拓片核对过。
18	#	著录格式—完全级 ISBD。
19	#	未定义,填空格。
20	4	每个地址目次款目中的“字段长度”—总为 4。
21	5	每个地址目次款目中的“起始字符位置”—总为 5。
22	0	“执行定义部分”的长度—总为 0。
23	#	未定义,填空格。

例 2 00816nam1#2200265###450#

本例表示的是一个丛帖新记录的头标。本记录的制作符合 ISBD 的规定,并与原帖核对过,还另制作有其中某一组成部分的分析记录,本记录为最高层记录。

字符位	值	注释
0—4	00816	整个记录的字符总数为 816 个,前两位字符填 0。
5	n	新记录。
6	a	记录类型—拓片。
7	m	书目级别—丛帖专著。
8	1	层次等级—最高层。
9	#	未定义,填空格。
10	2	指示符长度—总为 2。
11	2	子字段标识符长度—总为 2。
12—16	00265	记录头标和地址目次区的字符总数为 265 个,即数据基地址(数据紧接在目次区之后)为 265。