

G272.6

Z36

现代计算机档案著录标准化精要

张正强 主编

张正强 邹伟农 卞刚 编著

上海科学技术文献出版社

图书在版编目 (CIP) 数据

现代计算机档案著录标准化精要/张正强主编. —上海:上海科学技术文献出版社,2000.6

ISBN 7-5439-1609-6

I. 现... II. 张... III. 档案-著录格式-标准化-研究 IV. G272.6

中国版本图书馆 CIP 数据核字(2000)第 18286 号

责任编辑 何兰林

封面设计 何永平

现代计算机档案著录标准化精要

张正强 主编

*

上海科学技术文献出版社出版发行

(上海市武康路 2 号 邮政编码 200031)

全国新华书店经销

江苏常熟人民印刷厂印刷

*

开本 850×1168 1/32 印张 8.5 字数 243 000

2000 年 11 月第 1 版 2000 年 11 月第 1 次印刷

印数:1-1 100

ISBN 7-5439-1609-6/G·391

定价:44.00 元

第一章 现代计算机档案 著录的框架格式

现代档案计算机著录,因其涉及到档案目录数据库的建设质量,因而在国际档案界也引起了关注。因为,目录数据的质量是数据库建设的生命线,而数据质量的核心是数据的标准化与规范化,所以,在档案数据库的建设中,档案目录数据所构成的一条记录是这些数据中最重要的组成部分,它是档案部门各项业务工作的基础。因此,研究与探索档案目录数据格式的标准化与规范化就成为高质量地建设档案数据库的关键所在。

一、现代计算机档案著录格式

(一) 现代计算机档案著录格式的基本概念

在探索与研究现代计算机档案著录格式时,首先,我们必须对档案目录数据格式要有一个界定,即什么是档案目录数据格式。所谓的档案目录数据格式,通常指的是档案目录数据的组织形式。这种组织形式如果是对档案目录数据的逻辑关系而言的,则称为档案目录数据的逻辑格式;如果这种组织形式是档案目录数据的逻辑关系在计算机存贮器中的实现,则称为存贮格式。逻辑格式是从档案目录数据的逻辑关系上表述目录数据的,它与档案目录数据的实际存贮无关,是独立于计算机的;而存贮格式则是逻辑格式在计算机中的映射(即对应转换),是依赖于计算机的。

在此我们所讲的档案目录数据格式是档案目录数据的逻辑格式,即档案目录数据的著录格式。

(二) 现代计算机档案著录格式的两个层次

现代计算机档案著录格式的两个层次,是指档案目录数据的组织形式的两个层面,第一个层面是在宏观层面上,即对构成档案著录单位(在我国档案著录中通常也称之为著录对象,即被实际著录的一份文件或一个案卷)的目录数据的总体组织与安排,这种组织与安排构成一条目录记录的总体框架,从而形成档案目录数据著录的框架格式;第二个层面是在微观层面上,即对构成档案著录单位的目录数据的具体编码,这种具体编码构成一条目录记录的各数据项(或称数据元)在逻辑上的有序排列,从而形成档案目录数据的可被具体显示、操作、检索的执行格式。

由于现代计算机档案著录格式具有这样两个层次,显然,要探索与研究其著录格式的标准化,必须从这两个层次入手进行,从而通过对现代计算机档案著录的框架格式和执行格式的分析、比较与归纳,从中找出指导我国现代计算机档案著录的规律性认识。

二、档案著录的框架格式

(一) 手工档案著录的框架格式

1. 我国手工档案著录的框架格式的特点

为了更好地了解现代计算机档案著录的标准的框架格式,我们先从手工档案著录的标准的框架格式谈起。在手工目录中,对一份档案或一个案卷的著录结果,称之为条目,它对应于一张手工编制的目录卡片,其上著录的目录数据是按手工目录数据的著录格式组织的,这些目录数据包括:题名、责任者、密级与保管期限、载体形态、附注,等等。这些数据必须根据规范的目录数据著录格式的要求来安排。手工档案著录格式其标准化的形式是在档案著录规则中规定的,如以中华人民共和国行业标准《DA/T18-1999 档案著录规则》中所规定的手工档案目录数据的著录格式为例,对档

案著录单位的著录其标准格式如下所示：

例一：

分类号		档案馆代号
档号	电子文档号	缩 微 号
正题名 = 并列题名；副题名及说明题名文字；文件编号/责任者+附件。—稿本。—密级；保管期限。—时间。—载体类型；数量及单位；规格。—附注		
提要		
主题词或关键词		

例一所示的手工档案目录数据的著录格式，它在总体的组织与安排上是这样的，首先它把对著录单位所著录的目录数据分为排检与编号、著录正文、提要、主题词或关键词这四个部分，然后将这四个部分组织安排成四个段落，即第一段落安排“排检与编号”部分，第二段落安排“著录正文”部分，第三段落安排“提要”部分，第四段落安排“主题词或关键词”部分，于是就形成了手工档案目录数据著录的总体的框架格式——段落格式，如下所示：

例二：

排检与编号部分	(段落 1)
著录正文部分	(段落 2)
提要部分	(段落 3)
主题词或关键词部分	(段落 4)

根据以上著录的框架格式的安排，于是就形成了如下的手工档案著录的段落格式：

例三：

排检与编号部分	分类号	档案馆代号
	档 号	电子文档号 缩 微 号
著录正文部分	正题名 = 并列题名：副题名及说明题名文字：文件编号/责任者 + 附件。—稿本。—密级；保管期限。—时间。—载体类型：数量及单位：规格。—附注	
提要部分	提要	
主题词或关键词	主题词或关键词	

根据这一格式，一份“关于必须严肃处理党员干部中违法乱纪案件的通报”的文件，其著录结果如下：

例四：

C16052	00112
1-23-6-5	64.5
关于必须严肃处理党员干部中违法乱纪案件的通报 / 中央纪律检查委员会 + 关于陈寿阳接受外商贿赂、贪污公款的调查报告。—正本：通报。—秘密；永久。—1982.1.8。—11页；16开	
本文件通报了江苏省机械进出口公司少数人，接受外商贿赂，贪污公款，而受到开除党籍，并建议撤销行政职务，由司法机关依法惩办的事件。	
腐败分子 开除党籍 接受贿赂 贪污公款	

由上可知，我国手工档案著录格式的框架格式是段落式，这种段落式的框架格式是与我国国家标准《文献著录总则》及《国际标准书目著录（总则）》所规定的段落式的框架格式基本上相一致的，

也就是说,在国家标准《文献著录总则》与《国际标准书目著录(总则)》中,都规定了段落式为手工目录数据著录格式的框架格式,因此我国手工档案著录的段落式框架格式是标准化的。

这一著录格式还有一个特点,就是除了第一部分的排检编号项这一段落后,其他各段落根据档案目录数据的实际情况,每个数据项的著录都是可以回行的,也就是说每个数据著录项都是变长的,不是定长的。如正题名著录项,如果正题名比较短可以著录一行,如果正题名稍微长点也可以著录两行,再长些甚至可以著录三行、四行等,如下列格式所示:

例五:

TB552[12]	412403
2-4-31-12	75.67

天津市污染气象要素的研究 = Research of the Pollution Meteorological Element in Tianjing / 孙奕敏(天津市气象研究所副研究员);王长有,铁学熙(天津市气象研究所助理研究员) [等]. — 正本. — 永久. — 19821105—19851031. — 8卷. — 参加本课题的人员还有边海;李楦;陈英;刘学军. — 19881011 获得天津市科学技术进步二等奖

通过大气边界层中的湍流与气象特征来研究大气污染物质的输送和扩散稀释规律,为大气污染的预报提供了科学依据。

大气污染 气象要素 大气边界层

例五这一条目中,用黑体字显示的分别是目录数据的“题名”、“责任者”和“附注”,从该例中我们可以看出:由于该文件的题名与责任者项的字数比较多,所以可以回行,共占了三行;附注项内容的字数也比较多,因此也占了三行。由此我们可以看出这种段落格式的框架格式既有标准化、规范化的一面,又有适应档案目录数据实际的灵活性的一面。这种灵活性对比一下表格式的著录格式就可以看的更明显,如下所示:

例六：

分类号：TB55[12]		档案馆代号：412403
档 号：2-4-31-12		缩微号：75.67
题 名：天津市气象要素的研究=Reseach of the Polution Meteorologic		
责任者：王长有；铁学熙（天津市气象科研 文件编号：		
密 级：	时间：19821105-19821031	附注：参加本课题的人员
文 本：正本	载体类型标识：	载体形态：8 卷
主题词：大气污染 气象要素 大气边界层		

——完整的题名与责任者，附注项著录不下。

——由于表格式是定长，文号、密级与载体类型标识项内无数据，又浪费了空间。

从例六表格式对目录数据的处理我们可以看出，由于表格式对每一个目录数据项都给出了固定的著录空间，也就是说定长的空间，所以一旦著录空间容纳不了目录数据项的内容，也只好以牺牲内容为代价；反之，当数据著录项内无数据时，又浪费了空间。所以，这种著录格式不能适应档案目录数据不定长的实际情况，不是档案目录数据著录的标准的框架格式。

由此，我们可以总结出我国目前手工目录数据著录的标准的框架格式是：

可变长的段落格式

也就是这个格式统一、规范了我国档案目录数据著录的框架格式，而且在国际档案界，在这方面我国的标准化工作也是起步比较早的。

2. 我国手工档案著录的标准框架格式的意义

手工档案著录的框架格式必须是标准的、统一的。因为框架格式的标准与统一有利于建立高质量、高效率的检索体系,有利于建立全国统一的目录中心和编制联合目录,有利于馆际间的档案目录信息的交流,等等。如果没有这样的著录格式,档案目录数据的组织与安排就会出现无序,那么,档案目录数据既可以按如下所示的格式组织与排列:

例七:

关于必须严肃处理党员干部中违法乱纪案件的通报/中央纪律检查委员会。一正本。一秘密;永久。—1982.1.8。—11 页;16 开
+关于陈寿阳接受外商贿赂、贪污公款的调查报告

C16052

00112

1-23-6-5

64. 5

腐败分子 开除党籍 接受贿赂 贪污公款

本文件通报了江苏省机械进出口公司少数人,接受外商,贪污公款,受到开除党籍,建议撤销行政职务,由司法机关依法惩办的事件。

同时目录数据也可以按如下所示的格式组织与排列:

例八:

腐败分子 开除党籍 接受贿赂 贪污公款

本文件通报了江苏省机械进出口公司少数人,接受外商,贪污公款,受到开除党籍,建议撤销行政职务,由司法机关依法惩办的事件。关于必须严肃处理党员干部中违法乱纪案件的通报/中央纪律检查委员会。一正本。一秘密;永久。—1982.1.8。—11 页;16 开 +关于陈寿阳接受外商贿赂、贪污公款的调查报告

C16052

00112

1-23-6-5

64. 5

以至于还可以有各种其他安排。这样显然是不行的,因为这样,档案的条目就无法排序,条目无法排序就无法形成一个统一的检索体系,没有一个统一的检索体系也就无法对作为战略资源的档案信息进行有效的检索与利用,更谈不上档案目录中心的建立、档案目录信息的交换和档案信息资源共享的实现。所以档案目录数据著录的框架格式不实行标准化是不行的。

档案著录的框架格式究竟要达到什么样的标准化程度,这主要取决于在多大范围内的信息交换与资源共享,如果要在一个行业内实行档案信息的交换与资源共享,那么就应该在一个行业内实行档案目录数据著录的框架格式的标准化;如果要在一个国家内实现文献信息的交换与资源共享,那么就要在一个国家内实行目录数据著录框架格式的标准化;而如果要在国际上使整个人类社会的信息资源都能得到有效的资源配置,那么就要在国际范围内实行目录数据著录的框架格式的标准化。

显然,从人类社会的进步而言、从一个国家的战略信息资源的有效开发与利用而言,目录数据著录的框架格式必须实行国家标准化或国际标准化。这也就是目录数据著录的框架格式在一个国家范围内其行业标准必须与国家标准相兼容、在国际范围内必须与国际标准相兼容的原因。

所以,统一的档案著录的框架格式是信息资源共享的基础,没有统一的档案著录的框架格式,就不可能有信息资源的有效配置与有效利用。

(二) 现代计算机档案著录的框架格式

1. 现代计算机档案著录的标准框架格式的特点

从上面手工档案著录的标准框架格式的必要性及其重要意义的分析,我们很显然地可以逻辑地得出这样一个结论,那就是在现代计算机档案著录的情况下,必然也需要一个标准的框架格式,不能想象,在现代计算机就是网络的条件下,计算机的档案著录却没有一个标准的框架格式而可以随意地进行。

那么什么是计算机著录的框架格式呢? 所谓的计算机著录的框架格式就是对构成机读目录记录的目录数据的总体组织与安排。如前所述, 统一的框架格式有助于目录数据的交换与信息资源的共享, 所以, 计算机著录的框架格式又通常被称之为交换格式。那么在现代计算机档案著录中有没有可遵循的标准的框架格式呢? 是有的, 那就是《ISO-2709 1996 信息和文献工作——信息交换格式》(以下简称国际标准 ISO2709) 中所规定的格式。由于这一格式是文献目录数据著录通用的标准框架格式, 所以, 也就理所当然地成为现代计算机档案著录应遵循的标准框架格式。

为什么现代计算机档案著录要遵循这一格式, 其实际究竟有何意义呢? 下面我们就来分析一下。首先, 该标准的框架格式是把一条机读目录记录在总体上组织成四个区:

例九:

区	1
区	2
区	3
区	4

相对应于这四个区, 一条机读目录记录中的数据也就相应地被安排成四个部分:

例十:

序 号	部 分	所 在 区 位
第一部分	头标(即为一条记录的总体说明)	区 1(通常称头标区)
第二部分	目次(即为一条记录的数据目录)	区 2(通常称目次区)
第三部分	数据(即为一条记录的著录正文)	区 3(通常称数据区)
第四部分	记录分隔符	区 4

根据上面的总体组织与安排,由此形成了下列的计算机著录的框架格式:

例十一:

头	标	区
目	次	区
数	据	区
记	录	分 隔 符

目前,国际档案界在网络上的机读目录数据的组织与排序基本上都是按照这一个框架格式进行的,由于这一框架格式是国际标准化组织早在 1973 年就将其规定为标准的框架格式,并于 1996 年在《ISO—2709 1996 信息和文献工作——信息交换格式》(即 ISO2709 的第三版)中又再次将该格式规定为标准的框架格式,所以是高度标准化的。

我国早在 1982 年也参照了国际标准 ISO2709 第二版制定了与该国际标准相兼容的国家标准《GB 2901 文献目录信息交换用磁带格式》,并于 1992 年进行了修订,改为国家标准《GB/T 2901-92 书目信息交换用磁带格式》(以下简称国家标准 GB2901),也将该框架格式作为国家标准加以规定。

遵循这个标准的框架格式,其关键就是能使各计算机系统所生成的机读目录数据记录在进行数据交换时都有了一个标准格式,从而扫除了数据的共享和规模化利用中的障碍。从没有标准化就没有现代化的角度讲,是否按标准化的框架格式组织数据,是数据能否达到标准化及其高质量的标志之一。或着说,要实现信息资源的共享、要实现档案信息的网络化利用,不按标准化的框架格式组织数据是不行的。

下面就是一条按这一标准框架格式实际著录而生成的机读目录数据记录:

例十二：

- 1、头标区 → 00764 n b m 1 # 22001951##450# 001001100000
- 096002400011 100004100035 101000800076 200015700084 205000
- 900241 210001300250 215001400263 300002700277 430005100304
- 2、目次区 → 606003900355 606004700394 694001400441 711003500455 8010045
- 00490 905001700535@ 34-2804-34@##\$a [1955]国资工字第 041
- 号@##\$a19990901j195508200c#0chiy0121####ca@0#\$achi@0#\$a
- 湖北省人委国家资本主义办公室关于洪湖县七月十一日洪工字第
- 3、数据区 → 0 六号来函所提问题的答复@Sc[关于对洪湖县划分个体手工业
- 户、农业户的标准复文]@Sf 湖北省人委国家资本主义办公室
- @##\$a 正本@##\$a8 页@\$d16 开@##\$a 题名根据文件内容自拟
- @#0\$12000#\$a 关于洪湖县划分个体工业户、农业户的标准的
- 报告@##\$a 手工业\$х 社会主义改造\$у 洪湖县\$2cat@##\$a 资本
- 主义工商业\$х 社会主义改造\$у 洪湖县\$2cat@##\$aA9372\$у2@
- 02\$a 湖北省人委\$b 国家资本主义办公室@#1\$aCN\$b 上海空军
- 4、记录分隔符 政治学院信息管理系\$с19990511@##\$a442001@ %

(1) 记录头标区内容与作用的分析

根据国际标准 ISO2709 与国家标准 GB2901 的规定,头标区作为框架格式的第一部分位于各条记录之首,以例十二为例,一条根据标准框架格式实际生成的机读目录记录的头标区如下:

例十三:

0 0 7 6 4 n b m 1 # 2 2 0 0 1 9 5 1 # # 4 5 0 #

标准的机读目录记录,其头标区的长度均为 24 个字符,头标区中的代码是由字符位置标识的,其头标区中的代码与字符位的对应关系如下所示:

例十四:

代 码→0 0 7 6 4 n b m 1 # 2 2 0 0 1 9 5 1 # # 4 5 0 #

字符位→0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

为了便于理解头标区的内容,我们将其列成如下格式加以说明:

例十五:

字符位	代码	作 用	说 明
0	0	说明本记录的长度	代码“00764”表示该记录的长度为 764 个字符,由于国际标准规定记录长度最长可达 5 位十进制数,而本记录只有 3 位数,故前两位填“0”。
1	0		
2	7		
3	6		
4	4		
5	n	说明记录状态	“n”表示该记录是一条新记录,不是一条老记录。
6	b	说明记录类型	“b”表示该记录所反映的文献类型是档案。
7	m	说明著录等级	“m”表示该记录是文件级著录。
8	1	说明等级关系	“1”表示该记录在著录等级中是处于最高等级的记录。

(续表)

字符位	代码	作用	说明
9	#	说明档案控制	“#”表示是一空格,在该位置未定义代码。
10	2	说明指示符	“2”表示该记录数据字段所用的指示符长度是“2”。
11	2	说明子字段标识符	“2”表示该记录数据字段所用的子字段标识符长度是“2”。
12	0	说明数据 基地址	代码“00195”表示该记录的第一个数据字段相对于记录的起始位置(即数据基地址),由于国际标准规定数据基地址的长度为5位十进制数,而本记录只有3位数,故前两位填“0”。
13	0		
14	1		
15	9		
16	5		
17	1	说明著录级别	“1”表示记录在著录时,未与著录实体核对过。
18	#	说明著录格式	“#”表示完全采用中国档案著录规则。
19	#	目次说明	“#”表示是一空格,在该位置未定义代码。
20	4		“4”表示根据国际标准规定,字段长度设计为4位十进制数。
21	5		“5”表示根据国际标准规定,字段起始位置长度设计为5位十进制数。
22	0	说明执行定义	“0”执行定义部分的长度,总为“0”。
23	#	备用	“#”表示是一空格,未定义,留以后扩充用。

由上面对头标区各代码所表示的内容的分析,我们可以看出,标准框架格式的头标区实际上是对一条记录的总体说明,它主要说明了如下两个方面:

① 给出了一系列指示性数据:如记录长度、数据基地址、目次说明等,供计算机识别与处理记录时使用。

② 给出了一系列说明性数据:如记录状态、记录类型、著录等级、著录级别、著录格式、档案著录控制等,对记录本身起自我说明作用,同时也供计算机识别与处理用。

由此我们可以知道,记录头标区在机读目录中起着识别、说明

一条记录的作用,一条记录有了标准化的头标区,那么这条记录在任何计算机系统中都可以被识别、被处理。同时在时间上经过几十年、几百年,这条记录自身的总体情况依然是清晰的,可确认的。这就是为什么要在标准框架格式中设置头标区的意义所在。

(2) 记录目次区内容与作用的分析

根据国际标准 ISO2709 与国家标准 GB2901 的规定,目次区作为框架格式的第二部分是紧跟在头标区之后,其由若干目次和置于目次区末的字段分隔符组成。同样以例十二为例,一条根据标准框架格式实际生成的机读目录记录的目次区内容如下:

例十六:

```
0010011000000960024000111000041000351010008000762
0001570008420500090024121000130025021500140026330
0002700277430005100304606003900355606004700394694
001400441711003500455801004500490905001200535@
```

标准的目次区是由若干个目次项和置于区末的字段分隔符“@”构成,每个目次项对应一个数据字段。一个目次项由字段标识符、数据字段长度和数据字段起始位置三部分构成。一条记录中有几个数据字段,目次区中就有几个目次项。上面所示的目次区共有 16 个目次项,说明这条记录在数据区共有 16 个数据字段,具体如下所示:

例十七:

<u>001001100000</u> 目次项 1	<u>096002400011</u> 目次项 2	<u>100004100035</u> 目次项 3	<u>101000800076</u> 目次项 4
<u>200015700084</u> 目次项 5	<u>205000900241</u> 目次项 6	<u>210001300250</u> 目次项 7	<u>215001400263</u> 目次项 8
<u>300002700277</u> 目次项 9	<u>430005100304</u> 目次项 10	<u>606003900355</u> 目次项 11	<u>606004700394</u> 目次项 12

694001400441	711003500455	801004500490	905001200535
目次项 13	目次项 14	目次项 15	目次项 16

①

字段分隔符

根据国际标准 ISO2709 与国家标准 GB2901 的规定,每一个目次项均由 12 位字符构成,其中 1 至 3 字符位是用三位十进制数字表示的字段标识符,4 至 7 字符位是用四位十进制数字表示的字段长度,8 至 12 字符位是用五位十进制数字表示的相应的数据字段的起始地址。为了便于分析,我们将上面 16 个目次项列成下列格式,对其内容进行说明:

例十八:

目次项	字段标识符 (1—3 位)	字段长度 (4—7 位)	起始地址 (8—12 位)	说 明
1	001	0011	00000	001 标识符表示 001 这一数据字段的目次项,该数据字段长度为 11,其起始地址为 00000。
2	096	0024	00011	096 标识符表示 096 这一数据字段的目次项,该数据字段长度为 24,其起始地址为 00011。
3	100	0041	00035	100 标识符表示 100 这一数据字段的目次项,该数据字段长度为 41,其起始地址为 00035。
4	101	0008	00076	101 标识符表示 101 这一数据字段的目次项,该数据字段长度为 8,其起始地址为 00076。
5	200	0157	00084	200 标识符表示 200 这一数据字段的目次项,该数据字段长度为 157,其起始地址为 00084。