



中国高等教育文献保障系统

CALIS联机合作编目培训班

(中文图书)

2009年6月8-12日·辽宁大学图书馆



本讲义版权归CALIS联机编目中心所有，仅用于此次培训，任何单位及个人未经允许，不得翻印或用作它用！



第二讲

CNMARC格式介绍



内 容

❖ MARC格式介绍

- MARC是什么?
- CALIS采用的MARC格式

❖ CNMARC书目格式介绍

- CNMARC格式的结构
- 记录头标
- 目次区
- 数据字段
- 约定符号

❖ 机读目录和卡片目录的差异

MARC是什么？

- ❖ MARC是Machine Readable Catalogue的缩写形式，即机器可读目录。
- ❖ MARC格式（MARC Format）即机器可读目录格式。

MARC是什么？

- ❖ MARC格式由美国国会图书馆于20世纪60年代开始研制。
- ❖ 为了使计算机能识别目录，软件人员设计了MARC格式。
- ❖ 1966年2月美国国会图书馆成功试制了MARC I，开始发行MARC磁带。
- ❖ 经过几年不断试验和改进，研制出MARC II

MARC是什么？

- ❖ 1971年美国标准局正式批准MARC II 格式为美国国家标准
- ❖ 1973年国际标准化组织将MARC格式作为国际标准正式颁布，即所谓USMARC格式，或称LCMARC，现改名为MARC21
- ❖ 沿用欧美目录的一些国家以及机读目录发展比较早的国家，一般采用USMARC

MARC是什么？

- ❖ 继美国之后，又有许多国家陆续进行了本国机读目录系统的开发和研制
 - 1969年英国的UKMARC
 - 20世纪70年代德国、法国、加拿大、澳大利亚、意大利、比利时、挪威、丹麦、尼日利亚等国先后建立的本国机读目录系统。

MARC是什么？

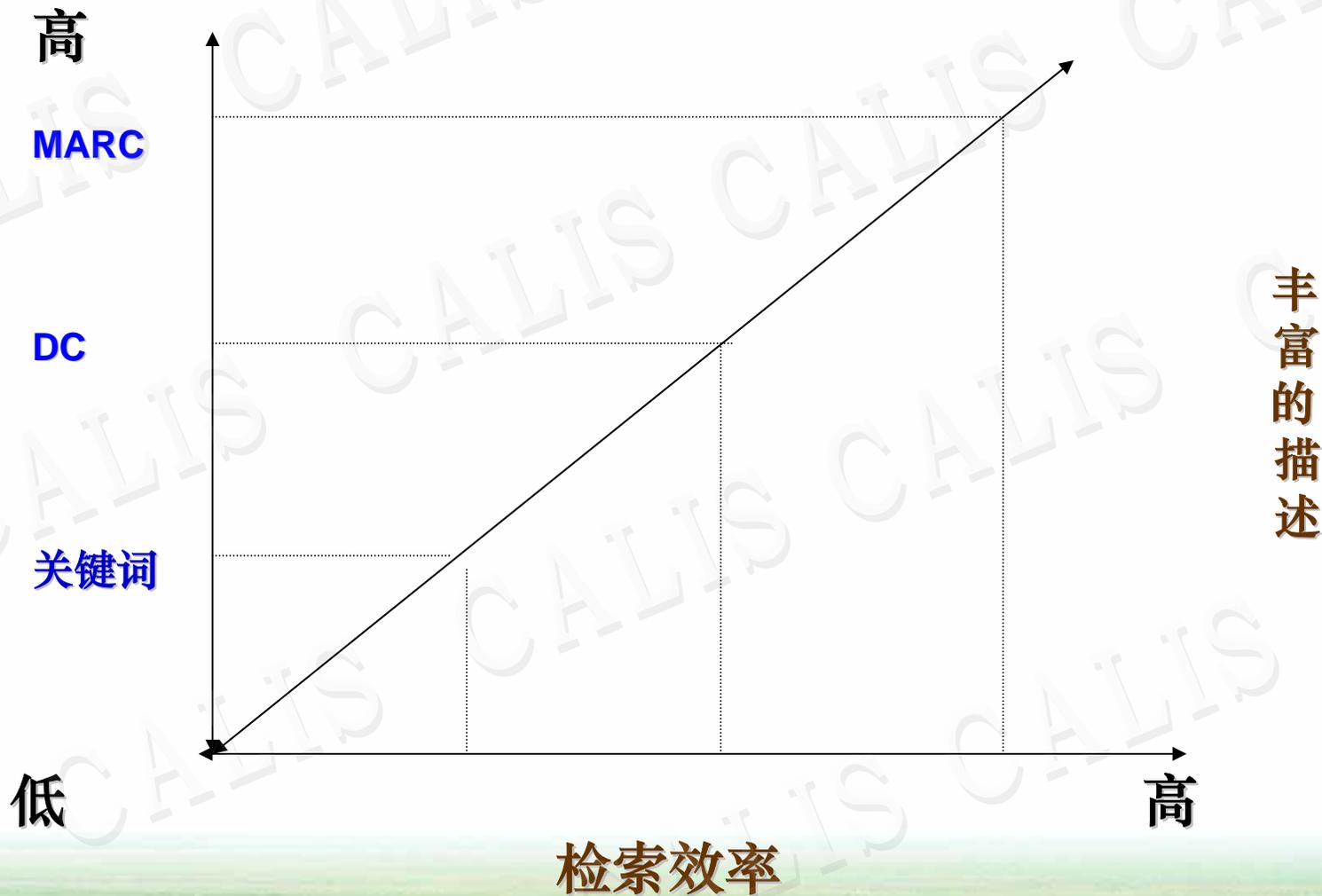
- ❖ 为了统一各国机读目录格式，国际图联（IFLA）于1977年主持研制UNIMARC，并于1994年出版了《UNIMARC手册》第二版。
- ❖ UNIMARC实现了不同文种、不同载体的文献机读目录格式的一体化，为不同国家书目机构之间机读目录的交换创造了条件，推动了各国机读目录格式的研制和修订。
- ❖ CNMARC是依据UNIMARC制定

MARC是什么？

- ❖ MARC是元数据（Metadata）的一种
 - 这里所指的元数据是一种关于数据结构化的数据
 - 是关于资源的信息
 - 是一种编目信息
- ❖ 元数据利用了知识管理的方法组织资源的信息，使用元数据可以提高知识使用的效益，或者说提高了信息被有效的检索和利用。
- ❖ 采用何种格式来描述资源的信息，取决于馆藏的规模、馆藏结构和馆藏文献的特点等因素。



丰富的描述与知识检索效应



MARC是什么？

- ❖ 丰富的描述与知识检索效应成正比
- ❖ 前页图表说明对资源的信息内容描述越详细，检索效益越高。
 - 关键词描述最简单，但是检索效益最低；
 - MARC描述最复杂，检索效益最高，但是成本也最高；
 - 基于DC的元数据介于其中。



CALIS所采用MARC格式

- ❖ CALIS联合目录数据库大多数印刷型文献采用MARC格式：
 - 基于UNIMARC的CNMARC格式
 - 中文、日文书目采用CNMARC格式
 - MARC21格式
 - 西文、俄文采用MARC21格式
- ❖ 采用不同格式的目的
 - 方便共享
 - 目前的技术能够实现不同格式的记录的一体化储存、一体化检索和一体化显示



CNMARC格式

- ❖ CNMARC格式是依据《UNIMARC手册》制定的
- ❖ 1986年北京图书馆自动化发展部依据UNIMARC格式编写成《中国机读目录通讯格式》
- ❖ 1992年由书目文献出版社正式出版
- ❖ 1996年7月1日作为文化部行业标准的《中国机读目录格式》正式实施，即CNMARC格式



CNMARC书目格式介绍

- ❖ CNMARC的书目格式提供了对各种文献类型进行著录、检索的详细说明
- ❖ 提供对书目控制所需要的各种编码数据元素的详细说明
- ❖ 是综合各种文献类型的一体化格式
- ❖ 利用该格式可以编制普通图书、连续性资源、测绘制图资料、可视资料、古籍、金石拓片、电子资源和乐谱等多种文献的书目记录



CNMARC书目格式介绍

❖ CALIS联合目录数据库中现有

- 图书 (Monograph)
- 连续出版物 (Serial)

❖ 计划发展

- 电子资源 (Electronic Resource, 原名计算机文件)
- 乐谱 (Music) 资料
- 录音资料 (Sound Recordings)
- 影片和录像资料 (Motion Pictures and Videorecordings)
- 地图 (Map)



CNMARC格式结构

MARC记录建立的目的是为了全球共享，共享的前提是数据的可交换性，MARC记录的交换格式是ISO 2709（GB 2901）。

头标	目次区	数据字段区	记录终止符
----	-----	-------	-------



CNMARC格式的结构

以下是一条完整书目记录的ISO 2709格式，仅在数据交换时采用这种格式*

记录头标

❖ 定义

- 头标位于每个记录的开头
- 固定长度为24个字符位置（00-23）
- 由一系列按字符位置定义的数字代码和字母代码组成，提供对记录进行处理的参数。

记录头标

❖ 特点

- 头标处于每一个记录的最开始位置，每一条MARC记录都必须具备，不可重复
- 头标中不包含字段标识符、字段指示符和子字段标识
- 固定长度为24个字符位置（00-23）。包括记录长度、记录状态、执行代码、指示符长度、子字段标识符长度、数据基地址、记录附加定义、目次区结构等数据元素。
- 头标的元素按字符位置定义

记录头标

❖ 头标数据元素列表 *

❖ 需要选择的数据元素：

- 记录状态 (05字符位)
- 记录类型 (06字符位)
- 书目级别 (07字符位)
- 层次等级代码 (08字符位)
- 编目等级 (17字符位)
- 著录格式 (18字符位)

记录头标

❖ 记录状态（05字符位）

➤ 反映书目记录的维护状态，记录状态定义了5种代码。CALIS联合目录中，选用3种：

➤ c 修改过的记录

表示修改过或更新过的记录。例如，对原有记录添加了信息或者对其中的一处或数处的数据作了修改。

➤ n 新记录

表示进入书目文档后一直没有被修改过的记录。新记录包括编目文献在出版前就编制的记录（如CIP在版编目记录），也包括几十年前就已建立，但进库后没有作过任何修改的记录，还包括简编记录。

➤ p 由在版编目升级的记录

表示由在版编目记录（CIP）升级而成的记录。

记录头标

👉 注意要点:

- ❖ CALIS联合目录规定不启用代码**d**（被删除的记录）和**o**（有较高层次的记录）。
- ❖ CALIS联合目录规定，只有对书目记录进行了较大的修改或更新（主要是检索点发生了变化）才采用代码**c**。
- ❖ 代码**p**仅用于套录在版编目记录时。

记录头标

❖ 记录类型（06字符位）

反映记录类型，定义了15种代码。适合图书编目的代码1种：

a 印刷型文字资料

包括印刷型文字资料及其缩微品，但不包括手稿型文字资料。

记录头标

❖ 书目级别(07字符位)

说明书目级别，定义4种代码。适合图书编目的代码有2种：

➤ m 专著

一册或多册按一定计划出版的资料，在图书编目范畴，通常指单行本专著或多卷本专著（多卷书）。

➤ s 连续出版物

以分册或卷期形式连续出版的资料，在图书编目范畴，通常指集中著录的丛书（编）、年鉴、工具书等。

记录头标

注意要点：

- 集中著录的多卷本专著的书目级别选用**m**，而不是s
- 分散著录的丛书的单册书目级别选用**m**，而不是s
- CALIS联合目录要求年鉴、年刊等具有连续出版物特性的文献按连续性资源编目。

记录头标

❖ 层次等级代码（08字符位）

说明本记录以层次性的关系与其它记录连接以及它在层次等级中的相对位置。具有层次等级关系的相关记录应共存在同一个数据库中。定义了4种代码。

记录头标

❖ CALIS联合目录规定仅用2种代码：

➤ # 层次等级关系未定

分散著录的丛书的单册，分散著录的多卷书采用代码“#”。

0 无层次的记录

无等级关系的单册图书、集中著录的多卷书采用代码“0”。

记录头标

注意要点：

- ▶ CALIS联合目录不要求做高层记录，所以不使用代码“1”和“2”。
- ▶ CALIS联合目录数据库中不建立高层记录，因此对于那些虽然有层次关系但未作连接的文献，无论是丛书的单册，还是分散著录的多卷书，均为“#”。

记录头标

- ❖ 编目等级（17字符位）说明书目记录的书目信息或内容标识的完整程度。
 - # 完全级
 - 指最完整的MARC记录。编制记录时与编目实体核对过。
 - 1 次级1(未核实文献实体)
 - 最完整的MARC记录，但在建立该记录时未与编目实体核对过。该级别主要用于回溯转换的记录，某些控制字段的代码和某些数据元素仅仅基于卡片目录的著录信息。

记录头标

➤ 2 次级2

- 该记录为在版编目（CIP）记录，这类记录通常是不完整的，如载体形态字段可能没有或不完整。

➤ 3 次级3

- 该记录属于不完整编目的记录，该记录有可能由发行机构升级为完全级的记录，也有可能不再被升级。

➤ 4 次级4

- CALIS联合目录自定义级别，该记录属于不完整的书目记录，通常由采访人员根据订单建立。

记录头标

👉 注意要点

- 简编记录编目级别=3
- 采访级简编记录编目级别=4
- 据规范的卡片目录回溯=1（名称标目/主题基本符合当前要求）
- 根据图书实体回溯，名称标目/主题完全符合当前要求，级别=#

记录头标

- ❖ 著录格式（18字符位）说明编制记录依据的著录规则，在著录200-225字段时是否依据国际标准书目著录（ISBD），定义3种代码。
 - # 完全采用ISBD格式
记录中出现的著录数据元素全部符合ISBD规定
 - i 部分的或不完全的ISBD格式
记录中出现的著录数据元素只有一部分符合ISBD规定。
 - n 非ISBD格式
记录中未出现符合ISBD规定的著录数据元素。



记录头标

👉 注意要点

- ❖ CALIS联合目录规定著录格式一律采用ISBD规则，该字符位通常为空位“#”。

目次区

❖ 定义

- 目次区是记录中每个可变长控制字段和可变长数据字段位置的索引，由计算机自动生成。各编目系统的机内格式不反映目次区。

目次区

❖ 特点

- 位于记录头标之后；
- 由若干个目次项构成；
- 每个项目长度=12，由**字段标识符、字段长度和字段起始字符位置**三个部分组成；
- 包含的数据项目数与该记录所包含的字段数是相等的。如果一条书目记录由15个字段组成，则该记录的目次区包含15个目次项
- 目次区以字段分隔符结束。一个目次区的总长度应为 $12 \times N$ （款目数） + 1（字段终止符）



目次区结构示意图

字段标识符 (3)	字段长度 (4)	字段起始字符 (5)	F/T
目次项1			目次项2	目次项N		终止符



目次区结构

001 0017 00000	005 0017 00017
010 0028 00034	099 0021 00062
100 0041 00083	101 0008 00124
102 0015 00132	105 0018 00147
106 0006 00165	200 0075 00171
205 0010 00246	210 0031 00256

数据字段

❖ 三个层次：

- 功能块、字段、子字段或数据元素
- 子字段由数据元素构成，若干子字段构成一个字段，若干字段构成一个功能块

❖ 两个部分

- 可变长控制字段
- 可变长数据字段

数据字段

❖ 可变长控制字段：**00X**

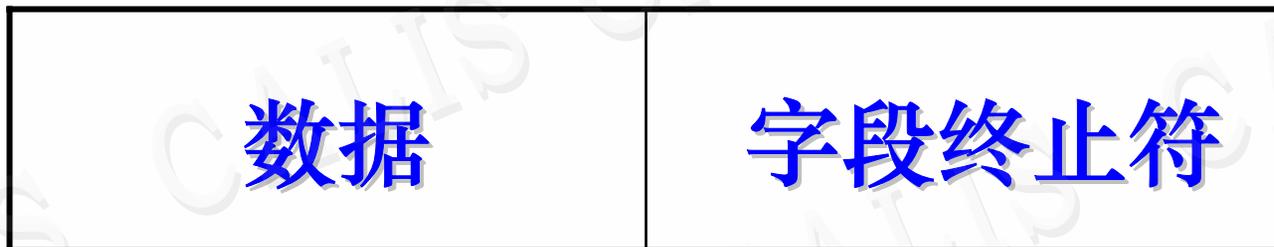
- 001 记录控制号
- 005 最近一次作业的日期和时间

➤ 特点

- 均无字段指示符和子字段代码；
- 由单个或多个数据元素组成；
- 通过字符位置识别相关数据元素；
- 大多数字段的长度为固定长度。



可变长控制字段结构



数据字段

❖ 可变长数据字段

➤ 字段标识符：01X-9XX

➤ 特点

- 均有字段指示符和子字段代码；
- 由单个或多个子字段组成，通过子字段标识相关数据元素；
- 均为可变长度。

可变长数据字段结构

指示符1	指示符2	子字段标识符	数据	...	字段终止符
1	#	\$a	北京故事	...	▲



数据字段 *

- ❖ CAL 012002005751 ▲ (001)
- ❖ 20020712093348.0 ▲ (005)
- ❖ ## ▼ -a7-111- 09335-6 ▼ dCNY68.00 (含光盘) ▲ (010)
- ❖ ##- ▼ a7-980007-34-4 ▼ -b光盘 ▲ (010)
- ❖ ## ▼ -aCAL 012002005751- ▼ aCAL 012002025931 ▲ (099)
- ❖ ## ▼ -a20020108d2002#####ekmy0chiy0121#####ea ▲ (100)
- ❖ 1# ▼ -achi- ▼ ceng ▲ (101)
- ❖ ## ▼ -aCN- ▼ b110000 ▲ (102)
- ❖ ## ▼ -aak##z###000yy ▲ (105)
- ❖ ## ▼ -ar ▲ (106)

.....



字段块

- ❖ 0XX 标识块
- ❖ 1XX 编码信息块
- ❖ 2XX 著录信息块
- ❖ 3XX 附注块
- ❖ 4XX 连接款目块
- ❖ 5XX 相关题名块
- ❖ 6XX 主题分析块
- ❖ 7XX 知识责任块
- ❖ 8XX 国际使用块
- ❖ 9XX 本地使用块

记录分隔符



位于每一机读记录结尾，用来区分记录的控制符号，通常标志本记录结束。

约定符号

	符号/代码	名称
字段/子字段/数据元素的必备性	M	必备 (Mandatory)
	A	有则必备 (Mandatory if applicable)
	O	可选项 (Optional)
字段/子字段重复性	R	可重复 (Repeat)
	NR	不可重复 (Never repeat)
字符代码	#	空位或未定义
	\$	子字段标识符中的分隔符

机读目录和卡片目录的差异

	机读目录	卡片目录
组成	一条记录可提供多种检索点	多种款目组成一条记录
信息量	信息量大	相对小
规范化	标准化规范化程度高	相对小
灵活性	多种功能：机读、卡片、书本、OPAC	单用途
时效性	制作周期短，使用时间长	制作周期长，使用时间短
共享性	可共享、可交换	不可共享、不可交换



本讲有关的常用名词术语*

名词术语

- ❖ **记录头标 (Leader)** : 简称头标。位于记录的开端, 固定长度为24位字符 (00-23), 其数据元素由一系列按字符位置定义的**数字代码**和**字母代码**组成, 提供本记录的有关参数。
- ❖ **目次区 (Directory)** : 由一系列固定长数据项目组成, 每个数据项目包含相应字段的字段标识符、字段长度和字段起始字符位置。

名词术语

- ❖ **记录数据元素 (Data element)** : MARC记录中被明确标识的最小信息单元。在可变长字段内, 数据元素由子字段标识符标识, 组成子字段; 在头标、目次区和定长字段内, 数据元素是按字符位置定义的数字或字母代码。
- ❖ **内容标识符 (Content designator)** : 用以识别MARC记录的数据元素, 或提供有关数据元素附加信息的编码。内容标识符包括字段标识符、指示符和子字段标识符。

名词术语

- ❖ **字段 (Field)** : 由字段标识符标识的被定义的字符串。在CNMARC格式中,除了00X字段,其它字段均包含一个或一个以上子字段。其长度包括字段指示符,子字段标识符,数据元素和字段分隔符。
- ❖ **字段标识符 (Tag)** : 用于标识字段的一组符号,由3位数字字符组成,贮存在目次区款目中。

名词术语

- ❖ **字段分隔符 (Field separator)** : 在每个可变长字段结束处采用的终止标识, 用以区分不同的字段。字段分隔符也用于目次区结束处, 它的标记为IS2 (ISO646的1/14) 。



名词术语

- ❖ **可变长字段 (Variable field)** : 长度不定的字段, 可包含一个或多个数据元素或子字段。字段的长度按字符数计算, 总长度为字段指示符、子字段标识符、字段分隔符以及文字数据所占的字符数的总和。
- ❖ **可变长控制字段 (Variable control field)** : CNMARC记录中的00X字段为可变长控制字段。可变长控制字段无字段指示符和子字段代码, 多数可变长控制字段由固定字符位的定长数据元素组成。
- ❖ **可变长数据字段 (Variable data field)** : CNMARC记录中的01X—9XX字段。可变长数据字段由指示符、子字段代码和长度不固定的子字段数据元素组成。

名词术语

- ❖ **指示符 (Indicator)** : 字段中头两个字符位的值。提供字段内容、字段之间的相互关系以及数据处理过程中所需操作的附加信息。每个指示符的值都有其独立的含义。字段指示符的值一般为数字或空位。
- ❖ **子字段 (Subfield)** : 字段内明确定义的最小数据单元。
- ❖ **子字段标识符 (Subfield identifier)** : 由两个字符组成, 用以识别可变长字段中的不同子字段。第1个字符为ISO 2709规定的专用符号, 标记为IS1 (ISO646的1/15), 第2个字符为字母或数字。



名词术语

- ❖ **定长子字段 (Fixed length subfield)** : 指在任何情况下长度不变的子字段。定长子字段可以包含一个或多个数据元素, 它可能是定长数据字段的子字段, 也可能是可变长字段的子字段。
- ❖ **填充符 (Fill character)** : 用于标识没有确切数据值的特定字符位置的字符, 而这个字符位置是必用的。 (|)
- ❖ **字符集 (Character set)** : 书目记录所需的常用字母、符号、发音符等字符及其录入、存储、显示方法的规定。常用的字符集有: ASCII、GBK、UNICODE等。

Thanks!

