

# 科技部科技基础性工作专项资金重大项目 研究成果

项目名称：我国数字图书馆标准规范建设

子项目名称：基本元数据标准规范

项目编号：2003DEA4T035

研究成果类型：应用指南

成果名称：基本元数据与 MARC 映射指南

成果编号：CDLS-S04-007

成果版本：总项目组推荐稿

成果提交日期：2005 年 12 月

撰写人：沈芸芸（北京大学图书馆）

## 项目版权声明

本报告研究工作属于科技部科技基础条件平台工作重大项目《数字图书馆标准与规范建设》的一部分，得到科技部科技基础条件平台专项资金资助，项目编号为2003DEA4T035。按照有关规定，国家和《我国数字图书馆标准规范建设》课题组拥有本报告的版权，依照《中华人民共和国著作权法》享有著作权。

为了学习、研究和应用推广等目的，可以复制、转载、或在电子信息系统上镜像本报告。但在复制、转载或镜像时，必须在明显地方标明“科技部科技基础条件平台工作重大项目《数字图书馆标准与规范建设》项目资助”的字样，必须保证本报告的完整性，必须注明总项目组、子项目组 and 作者的真实名称。任何人不得以商业赢利为目的复制、转载、镜像、或以其他形式传递和发布本报告。

报告版权人不承担用户在使用本作品内容时可能造成的任何实际或预计的损失。

## 作者声明

本报告作者谨保证本作品中出现的文字、图片、声音、剪辑和文后参考文献等内容的真实性和可靠性，愿按照《中华人民共和国著作权法》，承担本作品发布过程中的责任和义务。科技部有关管理机构对于本作品内容所引发的版权、署名权的异议、纠纷不承担任何责任。

《数字图书馆标准与规范建设》课题组网站(<http://cdls.nstl.gov.cn>)作为本报告的第一发表单位，并可向其他媒体推荐此作品。在不发生重复授权的前提下，报告撰写人保留将经过修改的项目成果向正式学术媒体直接投稿的权利。

# 基本元数据与 MARC 映射指南

## 目 录

1. 基本元数据核心集 .....	1
2. 基本元数据核心集与MARC互换的必要性.....	3
2.1 元数据描述数字资源的优势 .....	3
2.2 MARC描述数字资源的局限性 .....	4
2.3 元数据与MARC的关系 .....	5
3. 基本元数据核心集与MARC互换的可行性.....	5
4. 核心元素集与MARC的映射.....	7
4.1 核心集到MARC21 和CNMARC .....	7
4.2 CNMARC到核心集的映射 .....	15
5. 映射分析 .....	18
参考文献 .....	20

## 1. 基本元数据核心集

在科技部主持的我国数字图书馆标准规范建设项目的子项目“基本元数据标准”中，基本元数据核心集已经产生，这是一个资源描述的最小集。

表 1 基本元数据核心集

核心集元素	元素修饰词	编码体系修饰词
名称 (title)	交替名称 (alternative)	
创建者 (creator)		
主题 (subject)		LCSH (国会图书馆主题词表) [LCSH] MeSH (医学主题词表) [MeSH] CT (汉语主题词表) [CT] DDC (杜威十进分类法) [DDC] LCC (国会图书馆分类法) [LCC] UDC (国际十进分类法) [UDC] CLC (中国图书馆分类法) [CLC] LASC (中国科学院图书馆图书分类法) [LASC] FDC (四部分类法) [FDC]
描述 (description)	目次 (tableOfContents) 摘要 (abstract)	
出版者 (publisher)		
其他责任者 (contributor)		
日期 (date)	创建日期 (created) 生效日期 (valid) 可获得日期 (available) 发行日期 (issued) 修改日期 (modified) 接受日期 (dateAccepted) 赋予版权日期 (dateCopyrighted) 提交日期 (dateSubmitted)	Period (对于时间间隔的限定规范) [DCMIPeriod] W3CDTF (基于 ISO 8601 的规范的时间和日期的编码规则) [W3CDTF] EY (年号纪年, 以帝王在位期间用于纪年的名号为标准的阴阳历纪年法, 我国古代纪年法) [EY]

续表 1

核心集元素	元素修饰词	编码体系修饰词
类型 (type)		DCMIType (区分资源内容性质或种类的类型词汇表) [DCMIType]
格式 (format)	范围 (extent)	
	媒体 (medium)	IMT (因特网媒体类型) [MIME]
标识符 (identifier)		URI (统一资源标识符) [RFC2396]
	文献引用 (bibliographicCitation)	
来源 (source)		URI (统一资源标识符)
语种 (language)		ISO639-2 (国际标准化组织的语种识别代码) [ISO639] RFC3066 (因特网语种识别代码) [RFC3066]
关联 (relation)	版本继承 (isVersionOf) 版本关联 (hasVersion) 被替代 (isReplacedBy) 替代 (replaces) 被需求 (isRequiredBy) 需求 (requires) 组成 (isPartOf) 部分为 (hasPart) 被参照 (isReferencedBy) 参照 (references) 格式转换于 (isFormatOf) 格式转换为 (hasFormat) 遵循 (conformsTo)	URI (统一资源标识符)

续表 1

核心集元素	元素修饰词	编码体系修饰词
时空范围 (coverage)	空间范围 (spatial)	Point (DCMI 地理位置: 采用地理坐标点来标识空间的某一点) [DCMIPoint] ISO3166 (用于标识国家名称的 ISO3166 代码) [ISO3166] Box (DCMI 框图: 用地理界限来标识空间范围) [DCMIBox] TGN (地理名称词表) [TGN]
	时间范围 (temporal)	Period (对于时间间隔的限定规范) [DCMIPeriod] W3CDTF (基于 ISO 8601 的规范的时间和日期的编码规则) [W3CDTF] EY (年号纪年, 以帝王在位期间用于纪年的名号为标准的阴阳历纪年法, 我国古代纪年法) [EY]
权限 (rights)	访问权限 (accessRights)	

## 2. 基本元数据核心集与 MARC 互换的必要性

### 2.1 元数据描述数字资源的优势

元数据 (Metadata) 是指用于帮助识别、描述和定位网络化的电子资源的结构化数据。其主要目的就是通过准确地描述、评估信息资源来促进和提高信息检索, 这就明确了元数据的基本功能: 描述网络数据的内容; 使网络中的数据便于搜索, 有助于更准确地识别、定位和访问网络信息; 帮助用户决定某些数据是否为其所需; 等等。

核心集优势主要体现在以下方面:

- 实用性与灵活性。核心集基本是复用了 DC 的 15 个元素, 它是根据对数字信息资源的大量分析而总结出的结构化数据, 因而具有较强的实用性; 同时, 元素的使用不要求按严格的规则进行, 使用起来比较灵活。
- 元素的语义关系设计得比较清楚, 足以被广大的用户所理解, 而无须经过培训。

如元素“格式”的定义是“资源的物理形式或数字化表现形式”，元素“创建者”的定义是“对创建资源的内容负主要责任的实体”，等等，都十分简单明了，易于理解。

- DC 元素的类型易于确定，哪些元素是对内容特征的描述，哪些元素是对外部物理特征的描述，都是一目了然。

- 它不打算替代其它资源描述的标准，而是补充它们。它只打算描述电子文件的必要特性，以支持资源的发现。其它也很重要元素，如会计和档案数据，经过深思熟虑后都排除了，以确保格式尽可能的简单和可用。

- 基本上是句法独立的，以便应用于更广泛的范围和领域，所以句法上的连结是被竭力避免的。

- 所有元素都是可选的，但允许使用者定义哪些元素是必备的，哪些元素是任选的，这在很大程度上保证了可检索信息点的数量。

- 所有元素均可重复，如资源除正题名外可能有副题名、并列题名，允许重复元素“题名”及其子元素；资源可能有多个责任者，允许重复元素“其他责任者”，等等，这在很大程度上保证了资源描述的全面和深入。

- 通过特定限定词的使用，如用于元素“主题”中的主题词表的名称，使用者可以在有限范围内和已定义好的方式修改各元素。目前已定义好的可使用的主题词表有 LCSH、MeSH，各国还可根据本国的具体情况定义所需主题词表。

- 它是可以扩展的，以满足更多特定的社会需要。因为一些资源不可能通过一个小的元素集合就能够得到充分的描述。

## 2.2 MARC 描述数字资源的局限性

毋庸置疑，MARC 作为传统编目格式应用于数字信息描述必然会有其局限性。在文献资源由图书馆来组织和整理时，MARC 已经不可动摇地成为绝大多数图书馆采用的数据描述格式，它们足以应付馆藏资源的组织，但网络发展起来后，信息组织与检索发生了极大的改变，原有的书目描述无法跟上数字信息发展的需要了。MARC 的局限性主要表现在以下几个方面：

- ① 这种描述手段往往只适用于图书馆，除了图书馆，其它机构很少使用它。
- ② MARC 需要在专门的软件系统中使用，而且不太适应互联网的环境。在目前，图书馆自动化系统的发展，使 MARC 愈来愈受到系统的束缚，而这些系统之间又无法完全兼容，无法直接为用户提供全文信息，使信息的检索和传递受到很大的限制；同时，网络标记语言环境及无缝链接使得 MARC 无法马上融入到其中，网络信息的海量、多样及不稳定性也使它的弱点暴露无遗。

- ③ MARC 格式因体系庞大复杂，修订程序相当复杂，而且也非常缓慢。
- ④ MARC 适用于完整的、静止的信息内容的处理，不易处理动态的多媒体信息。

MARC 格式在设计之初还无法预料到网络信息时代的到来及其极大的影响，所以，它只考虑了当时有的文献类型——主要是印刷型文献，并预测到文献类型将会不断地发展变化，为格式的发展也准备了一定的空间。现在 MARC 及其相关标准的修订、增加 856 字段以及 ISBD (ER) 的出版，也证明了 MARC 并不是不能处理电子资源，但无疑，它对于传统的文献资料的过多的适应性，使其丧失了许多处理动态多媒体信息的能力。
- ⑤ 编制一条机读记录不仅需要经过严格的专业训练，而且需要花费相当的时间，这是不争的事实，也是困扰了图书馆界很多年的问题。

## 2.3 元数据与 MARC 的关系

元数据描述资源的优势显而易见，但是，它不是用于替代传统的资源描述方式的，虽然它是适应了网络化时代信息量激增、技术发展的需要，比传统资源描述的方式与手段更简洁有效。但 MARC 作为图书馆描述资源的元数据已经发展了数十年，不但积累了丰富的编目知识、经验以及共享的联机合作的体系，并且积累了数千万条的详细描述资源的记录，这是一笔巨大的财富，任何机构任何人都不能忽视，元数据的发展或多或少正在接受着传统编目的一些经验就是一项有利的证明。元数据的发展并不是要建立一个统一的、包罗的数据格式，而是要建立一系列的开放标准规范，在技术底层应用这些规范，以保证不同元数据间的互操作性。因此，元数据的设计应遵循一定的开放标准规范，对已有的元数据应建立彼此间的映射关系，使元数据间具有互操作性，使用户能够在开放的分布环境中发现、检索和利用所需要的任意资源和服务。

从以上分析可以看出，建立元数据与 MARC 间的映射是必要的。

## 3. 基本元数据核心集与 MARC 互换的可行性

因为元数据最大的功能就是用于信息检索，所以只有对检索有用的元素才具有标识的意义。在核心集中，这类元素可以被视为“类文献对象”（document-like objects, DLOs）。顾名思义，“类文献对象”就是“像文献一样的对象”，这与传统文献中书名页、封面、版权页等标明了诸多有用的可标识信息一样，网络中用它来表示具有类似于这些可标识信息的电子资源。一般认为，DLOs 是 Internet 资源检索中最通用的类型，任何处理 DLOs 的使用方案，都应能够推广至其它类型的资源。DLOs 没有被严格定义，但显而易见，网络环境下的 DLOs 包含着图像、声音与其它超文本信息，所以更复杂。DLO 的智力内容主要在于文本，这样，用于描述 DLOs 的元数据与传统印刷型

文本的元数据具有相似之处。如果对超出 DLOs 的对象描述所需要的元素不加以考虑，则在文本的描述上元数据间的差别是不大的。

核心集与 MARC 描述信息资源的特点的相似之处在于：（1）都集中于描述对象的内在特性，即通过被描述的文献对象可直接发现的特性（例如知识内容和物理形式方面的特性——题名、主题等）来揭示信息对象；（2）都可独立于描述对象而存在，尽管 DC 也可以与嵌入的 META 标记与 HTML 文本成为一体；（3）一经建立便可共享，而且在使用过程中可由各种人员为其增加新的属性，即逐步完善；（4）二者均可在技术的支持下，通过运用目录协议（如 whois++）、搜索和检索协议（如 Z39.50）、Harvest 等多种方式进行开发和利用；（5）DC 限定词的利用更加方便了二者映射的建立。

更重要的是元数据之间，它们与 MARC 之间都具有一些共同的属性，这些共同的属性对于建立它们之间的映射具有很大帮助。核心集与 MARC 共同的属性包括：

- 都包含有与数据分离的元素，只是名称不同。如在核心集中称元素，但在 MARC 中则称字段（field）、标识（Label）等。
- 每一元素都包含语义定义，即元素的名称、标签、定义、注释等，这样就可以找到其中相同或等效之处，如 TITLE 在元数据中表示资源的正式题名，TITLE.Alternative 表示交替题名（其他形式的题名），而 MARC21 中则用字段 245 表示资源的正题名，246 表示变异题名，等等。
- 元素是必备的、可选的或是根据特定的条件是必备的，都有明确的说明。一般在元数据中，各个元素及子元素都是可选用并可重复，但出于需要也可以规定某些元素是必备的。在 MARC 中，各个字段的属性也都是明确而且严格地规定好的。
- 元素间彼此关联，因而都受到相关元素在组织上的限制，如等级关系。如核心集在制定限定词时就有“向上兼容”的原则，每一限定词都有唯一的父族。MARC 中的等级关系更是严格规定，随处可见。
- 都有强加于元素的值中的限制，如各种规范的限制，在 MARC 中更有受控词表的限制。
- 对本地定义的元素的支持都是可以选择的，例如 CNMARC（基于 UNIMARC）中的 69X 字段<sup>①</sup>，UNIMARC 可以选择支持它们也可以选择不支持。

正是因为元数据与 MARC 间的这些描述信息资源的共同属性，它们之间的映射才有可能建立。核心集可能以不同的方式与 MARC 记录相互作用：（1）简单资源描述记录的提升。编目机构可能希望从以 DC 等格式（可假定为在 HTML 或 SGML 中的描述）提供的元数据中抽取数据并将其转换到 MARC 字段中，形成一条记录的骨架。编目人

---

<sup>①</sup> 690—中国图书馆分类法分类号；692—中国科学院图书馆图书分类法分类号；694 中国人民大学图书馆图书分类

员可能继续提升该记录以符合特定编目格式的需要。(2) 在句法和数据库中漫游检索。图书馆拥有以MARC书目记录信息为中心的庞大系统。随着近年来Internet上电子资源的增加, 其他句法结构也已被采用以提供元数据。美国国会图书馆与许多SGML专家合作为MARC创建了一个DTD<sup>②</sup>, 以便于转换能够在SGML和MARC之间以标准的方式实现。对系统技术来说, 能够以不同的句法在数据库中检索元数据以及使元素的定义与利用具有共性, 这是十分重要的。

基于以上分析, 核心集与MARC是可以相互转换的。二者的转换就是要在图书馆OPAC上能够检索到以DC著录的网络资源, 或者在DC检索界面上能够检索到以MARC著录的资源, 简言之, 即是在同一检索界面上能够使用户检索到不同数据库中的资源, 而不再需要进行数据库的切换, 并且能够达到较高的查准率。

## 4. 核心元素集与MARC的映射

### 4.1 核心集到MARC21和CNMARC

根据核心集与MARC的元素解释与语义说明, 可以建立映射如表2, 并在表后给出定长字段的赋值:

表2 核心集与MARC的元素解释与语义说明

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
名称	无	无	第一	245 00 \$a	200 1# \$a
	无	无	其他	246 33 \$a	517 1# \$a
	交替名称	无	任何	246 33 \$a	517 1# \$a

---

法; 等等。

<sup>②</sup> DTD实际上可以看作一个或多个XML文件的模板, 这些XML文件中的元素、元素的属性、元素的排列方式/顺序、元素能够包含的内容等, 都必须符合DTD中的定义。想要创建一份完整性高、适应性广的DTD是非常困难的, 因为各行各业都有他们自己的行业特点, 所以DTD通常是以某种应用领域为定义的范围, 如: 医学、建筑、工商、行政。

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
创建者	无	无	第一	100 1# \$a 110 2# \$a 111 2# \$a	701 #1 \$a 711 02 \$a 711 12 \$a
			其他	700 1# \$a 710 2# \$a 711 2# \$a	701 #1 \$a 711 02 \$a 711 12 \$a
			任何	720 ## \$a	730 0# \$a

注：由于创建者的情况在 MARC21 中有较为复杂的规定，因此无法判断到底是个人、团体还是会议责任者，如果不能判断则全部简化对应至 720 字段。

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
主题	无	无	任何	653 ## \$a	610 0# \$a
	无	DDC	任何	082 0# \$a	676 ## \$a
	无	LCC	任何	050 #4 \$a	680 ## \$a
	无	UDC	任何	080 ## \$a	675 ## \$a
	无	LCSH	任何	650 #0 \$a	606 0# \$a+2
	无	MeSH	任何	650 #2 \$a	606 0# \$a+2
	无	CT	任何	650 #7 \$a+2	606 0# \$a+2
	无	CLC	任何	093 ## \$a	690 ## \$a
	无	SKC	任何	650 #7 \$a+2	696 ## \$a+2

注 1：若所用主题为地区性的，则要在 \$2 子字段中注明所用的体系名称

注 2：若不说明主题词的来源，则 650 字段无需使用 \$2 子字段，并将指示符 2 的值改为 4

注 3：由于 MARC21 中主题的著录比较复杂，这里无法区分对待，如个人名称主题、团体名称主题、地理名称主题等等，只能笼统地对应到论题性主题 650 字段，若简化处理，则所有对应到 650 字段的主题均可直接对应到 653 字段。

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
描述	无	无	任何	500 ## \$a	300 ## \$a
	文摘	无	任何	520 3# \$a	330 1# \$a
	目次	无	任何	505 0# \$a	327 ## \$a

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
出版者	无	无	任何	260 ## \$b	210 ## \$c

续表 2

元素名称	元素修饰词	编 码 体 系	出现位置	MARC21	CNMARC
其他责任者	无	无	任何	720 ## \$a	730 0# \$a

注：同创建者注。

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
日期	无	任何	任何	260 ## \$c	210 ## \$d
	发行日期	任何	任何	260 ## \$c	210 ## \$d
	创建日期	任何	任何	518 ## \$a	300 ## \$a
	可获得日期	任何	任何	307 8# \$a	300 ## \$a
	生效日期	任何	任何	518 ## \$a	300 ## \$a
	修改日期	任何	任何	518 ## \$a	300 ## \$a
	接受日期	任何	任何	518 ## \$a	300 ## \$a
	赋予版权日期	任何	任何	260 ## \$c	300 ## \$a
	提交日期	任何	任何	518 ## \$a	300 ## \$a
		W3CDTF	任何		
		Period	任何		

注：对应到附注字段的部分应生成导语，如“生效日期：”、“提交日期：”等等

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
类型	无	任何	任何	516 ## \$a	336 ## \$a
	无	DCMIType	任何	245 00 \$h	200 1# \$b

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
格式	无	无	任何	516 ## \$a	337 ## \$a
	范围	无	任何	300 ## \$a	215 ## \$a
	媒体	无	任何	340 ## \$a	307 ## \$a
		IMT	任何	856 ## \$q	856 4# \$q

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
标识符	无	无	任何	024 8# \$a	001
	书目引文	无	任何	024 8# \$a	300 ## \$a
	无	URI	任何	856 40 \$u	856 4# \$u

注 1: URI 是统一资源标识符, 它可以进一步划分为定位符、名称或二者兼有。统一资源定位符 (URL) 和统一资源名称 (URN) 都是指 URI 的一个子集。

注 2: 024 的指示符为 8#, 可在 \$2 子字段中给出编码体系值, 如 ISBN, ISSN 等

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
来源	无	无	任何	786 0# \$n	324 ## \$a
	无	URI	任何	786 0# \$o	324 ## \$a

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21 字 段	CNMARC
语种	无	无	任何	546 ## \$a	302 ## \$a
	无	RFC 3066	任何	546 ## \$a+\$b	101 0# \$a
	无	ISO369-2	任何	041 0# \$a	101 0# \$a

注: 546 中的 \$b 指出编码体系为 RFC 3066。

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
关联	无	无	任何	787 0# \$n	488 #1 \$12001#\$a
	无	URI	任何	787 0# \$o	
	版本继承	无	任何	775 0# \$n	305 ## \$a
		URI	任何	775 0# \$o	
	版本关联	无	任何	775 0# \$n	452 #1 \$12001#\$a
		URI	任何	775 0# \$o	
	被替代	无	任何	785 0# \$n	442 #1 \$12001#\$a
		URI	任何	785 0# \$o	
	替代	无	任何	780 0# \$n	432 #1 \$12001#\$a
		URI	任何	780 0# \$o	
	格式转换于	无	任何	776 0# \$n	452 #1 \$12001#\$a
		URI	任何	776 0# \$o	
	格式转换为	无	任何	776 0# \$n	452 #1 \$12001#\$a
		URI	任何	776 0# \$o	
	被需求	无	任何	787 0# \$n	488 #1 \$12001#\$a
	需求	无	任何	787 0# \$n	488 #1 \$12001#\$a
		无	任何	538 ## \$a	
	组成	无	任何	773 0# \$n	461 #1 \$12001#\$a
		URI	任何	773 0# \$o	
	部分为	无	任何	774 0# \$n	462 #1 \$12001#\$a
		URI	任何	774 0# \$o	
	被参照	无	任何	510 0# \$a	321 ## \$a

		URI	任何	787 0# \$o	
	参照	无	任何	504 0# \$a	322 ## \$a
		URI	任何	787 0# \$o	
	遵循	无	任何	500 ## \$a	488 #1 \$12001#\$a

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21	CNMARC
时空范围	无	无	任何	500 ## \$a	300 ## \$a
	空间范围	任何	任何	522 ## \$a	607 ## \$a
	时间范围	任何	任何	513 ## \$b	122 2# \$a

续表 2

元素名称	元素修饰词	编码体系	出现位置	MARC21 字段	CNMARC
权限	无	无	任何	540 ## \$a	300 ## \$a
	访问权限	无	任何	506 ## \$a	337 ## \$a

MARC 格式远不只是这些字段的对应，还有头标区、定长字段等编码信息的值需要填写。

#### 1) MARC21 的头标区与定长字段

- 头标区包括 24 个字符位 (00-23) 的头标区是一个定长字段，在记录处理上应该根据元素中的值自动生成。
  - 06 字符位 (记录类型) 可以根据元素 Type.DCMIType 中的值自动生成 (见表 3):

表 3 根据元素 Type.DCMIType 中的值自动生成的 06 字符位

Type 中的值	06 字符位的值
Collection	p (混合型资料)
Dataset	m (计算机文件)
event	r (三维制品或天然物体)
image	k (二维非放映图像)
model	r (三维制品或天然物体)
party	r (三维制品或天然物体)
place	r (三维制品或天然物体)
service	m (计算机文件)
software	m (计算机文件)
sound	i (非音乐录音资料)
text	a (文字资料)

如果元素 Type 中没有值，则 06 位用 a。如果 Type 中有多个值，且其中之一为“collection”，则利用其它值来确定 06 字符位的值。

- 07 字符位 (书目级别): 若 Type 中的值为“collection”，则使用值“c”(合集); 若 Type 中的值为其它任何值，则使用值“m”(专著)。
- 08 字符位 (控制类型): 生成值“#”(空格: 无特定的控制类型)
- 09 字符位 (字符编码系统): 生成值“#”(空格: MARC-8)
- 17 字符位 (编目等级): 使用值“3”(简编级)，或可以使用的其它值
- 18 字符位 (著录标准): 使用值“u”(标准不详)，表明编目记录的著录标准无法确定

其它字符位的值可以预置，也可以自动生成，与普通 MARC 格式并无不同。

- 008 定长数据元素包括 40 个字符位 (00-39)，以编码信息的形式反映受编文献一般的及特殊的书目特征，用于检索和数据管理。对于可以从元素中生成的有关字符位的值如下：

- 00-05 记录入档日期或转换日期，一般由系统自动生成，形式为 YYMMDD
- 06 出版日期 / 状态类型，根据 DC.Date 中的值确定，若日期不清楚则为 n，同时 07—10 (日期 1) 和 11—14 (日期 2) 的值分别为 uuuu 和 uuuu；若为详细的日期 (即年月日均有) 则代码为 e，日期 1 中记录年份，日期 2 中记录月份和日期；若日期不同则为代码 m，并将年份分别记录在日期 1 和日期 2 中；若为单个已知或可能的日期则为代码 s，将单个的年份记录在日期 1 中，日期 2 则为空。

- 15-17 出版地代码, 如果 DC 中有提供 Publisher.Place, 则可经过 MARC21 国家代码表转换后形成

- 35-37 语种代码, 可以由元素 Language 在编码体系为 ISO639-2 或 MARC21-lang 时生成

- 其它字符位可以自动由计算机生成缺省值

此外, 一些可选用字段如 006 (定长数据元素—附件特征—一般信息) 和 007 (载体形态定长字段) 字段等, 因确定它们的值很难, 也不是必备字段, 故不再一一列出。

## 2) CNMARC 的头标区与定长字段

- 头标区包括 24 个字符位 (00-23) 的头标区是一个定长字段, 在记录处理上应该根据元素中的值自动生成。

- 00-04 字符位, 记录长度, 由系统根据记录的字符长度自动生成

- 05 字符位, 记录状态, 预置为 n

- 06 字符位 (记录类型) 可以根据元素 “类型” 中的值自动生成 (见表 4):

表 4 根据元素 “类型” 中的值自动生成的 06 字符位

Type 中的值	06 字符位的值
拓片	u (拓片)
古籍	v (善本)
舆图	v (善本)
地方志	v (善本)
家谱	v (善本)
音频资料	j (音乐录音资料)
视频资料	g (放映和视频资料)
电子图书	l (电子资源)
网络资源	l (电子资源)
电子期刊	l (电子资源)
学位论文	a (文字资料)
期刊论文	a (文字资料)
会议论文	a (文字资料)
课件	m (多媒体)
图像资料	k (二维图形)

- 07 字符位 (书目级别): 生成值 “m” (专著)。

- 08 字符位 (层次等级代码): 生成值 “#” (空格: 层次未定)

- 09 字符位 (字符编码系统): 生成值 “#” (未定义)
- 10-16 字符位 (指示符长度、子字段标识符长度、数据基地址): 系统自动生成
- 17-19 字符位 (记录附加定义): 生成值 “3i#”
- 20-23 字符位 (目次区结构): 生成值 “450#”
- 目次区: 自动生成
- 00 (记录控制号): 自动生成, 最好根据记录控制号生成规则
- 100 字段(共 36 个字符位):
  - 100 ## \$aYYYYMMDDdYYYY####ekmy0chiy0121####zz
  - 蓝字为自动生成, 预置 “d”, YYYY 为出版日期, 预置 “####ekmy0y0121####zz”
- 101 字段 (文献语种): 根据元素 “语种” 生成, 如 “中文”、“英文” 分别生成:
  - 101 0# \$achi
  - 101 0# \$aeng
 其他编码信息字段忽略。

## 4.2 CNMARC 到核心集的映射

以 MARC 为主到核心集的映射表是一项庞大的工程, 数以千记的字段与子字段要映射到一个只有几十个元素和修饰词的结构上, 结果是难以想像的。因为上面的映射表无法满足反向对应, 因此, 建立 MARC 到核心集的映射是必要的, 目前, 只做 CNMARC 常用字段到核心集的映射。

注 1: <> 中为核心集中没有但较常用的元素和修饰词

注 2: X 表示数字 0-9

注 3: 所有未列出字段与子字段可忽略

注 4: 多个字段或子字段对应于一个元素或修饰词时, 重复该元素或修饰词, 但 600-610 以及 7XX 字段除外, 以 “,” 号间隔各子字段

表 5 CNMARC 到核心集的映射

CNMARC 字段	子字段	核心集元素	修饰词	
			元素	编码体系
头标 <sup>[注 1]</sup>	字符位 6	类型		
001	无	标识符		
010	\$a	标识符		<ISBN>
011	\$a	标识符		<ISSN>
012-09X	\$a	标识符		
101	\$a { 多个则重复语种 }	语种		ISO639-2
200	\$a	名称		
	\$c	名称		
	\$d	名称	交替名称	
	\$e	名称	交替名称	
	\$z	名称	交替名称	语种
205	\$a, \$b	<版本>		
	\$f, \$g	其他责任者		
206,207,208	\$a	描述		
210	\$a	出版地		
	\$c	出版者		
	\$d	日期	发行日期	
215	\$a, \$c, \$d	描述		
225	\$a	关联		
230	\$a	描述		
30X	\$a	描述		
31X	\$a	描述		
320-323,325-326	\$a	描述		
324	\$a	来源		
327	\$a	描述	目次	
328	\$a	描述		
330	\$a	描述	文摘	
332	\$a	描述		

续表 5

CNMARC 字段	子字段	核心集元素	修饰词	
			元素	编码体系
333	\$a	<用户对象 audience>		
336	\$a	类型		
337	\$a	描述		
421, 422,423	\$12001#\$a	关联		
430,431	\$12001#\$a	关联		
432,433	\$12001#\$a	关联	替代	
434—441	\$12001#\$a	关联		
442,443	\$12001#\$a	关联	被替代	
444—448	\$12001#\$a	关联		
451, 452	\$12001#\$a	关联	版本关联	
461	\$12001#\$a	关联	组成	
462	\$12001#\$a	关联	部分为	
463—482	\$12001#\$a	关联		
488	\$12001#\$a	关联		
500, 501	\$a	名称	交替名称	
512—541	\$a	名称	交替名称	
600,601,602	\$a-\$h	主题		
606	\$a-\$z	主题	CT	
607	\$a	时空范围	空间范围	
608-610	\$a	主题		
690	\$a	主题	CLC	
701,711,721	\$a-\$h	创建者		
702,712,722,730	\$a-\$h	其他责任者		
856	\$u	标识符	URI	
856	\$q	格式	IMT	

注 1: 元素“类型”中的值根据头标 06 字符位给出, 06 字符位的字母意义如下, “类型”元素中直接写出文字, 而不是字母。

- a = 文字资料印刷品
- b = 文字资料手稿
- c = 乐谱印刷品
- d = 乐谱手稿
- e = 测绘制图资料印刷品
- f = 测绘制图资料手稿
- g = 录像制品、投影制品、电影制品
- i = 录音制品（非音乐）
- j = 录音制品（音乐）
- k = 二维图形（图画、设计图等）
- l = 电子资源
- m = 多载体
- r = 三维制品和教具
- u = 拓片

## 5. 映射分析

以上映射表是基于核心集和 MARC 格式的解释和说明形成的，这里面的对应是有歧义的，有些对应是牵强的，也可以说是不确定的。

毫无疑问，在理论上，任何结构化的数据都可以转换到另一种结构，但都不可能是完全对应的，都会产生一些数据的丢失。很多机构和个人都对它们之间的转换问题进行了研究，包括美国国会图书馆，普遍认为，不借助于人工干预的自动转换还存在着很多问题。

首先，MARC 是体系完备、内容及句法复杂的标准，而核心集目前仅有 15 个元素，加上一些修饰词，其所表达的意义与 MARC 表述丰富的字段、子字段相比相差很远。经过分析映射表，核心集元素与 MARC 字段、子字段之间存在着四种对应关系：

- ① 一对一关系。如：权限元素在 MARC21 中与 540\$a 相对应，出版者元素在 MARC21 中与 260\$b 相对应。
- ② 一对多关系。主题元素可以与 MARC21 中的多个不同的 6XX 字段相对应，因为在 MARC21 中，主题字段是分很多类型的，如 650 是普通论题主题，651 是地理名称主题，600 是个人名称主题；语种元素与 546（语种附注）、041（语种代码）以及 008 / 35—37 字符位对应；等等。
- ③ 多对一关系。如关联元素中不同的修饰词可以对应同一个字段，修饰词

IsFormatOf 与 HasFormatOf 都对应到 776（其他载体形态款目）字段。又如，格式和类型元素都可能对应到 516（电子文件类型或数据附注）字段中。

- ④ 无对应关系。有些修饰词是找不到恰当的对对应关系的，如果强制对应的的话，就只能用附注字段来表示，如关联元素中的遵循在 MARC 中无法找到恰当的对对应关系，故只能对应到 500 字段。

MARC 的结构复杂，基本分为头标区、目次区、数据区及记录分隔符四大部分，在著录时对受编文献的细节考虑较多，按照著录规定必须具备的字段与子字段较多，字段的共现要求也较多，而 DC 中则不加考虑，亦即 MARC 要求的许多必备元素是核心元素集的外部数据，故而转换起来难度很大。如受编文献是否是政府出版物、会议录、其价钱等。

其次，元数据讲求元素的句法独立性，其元素的意义不依赖于句法结构；而 MARC 对字段的句法要求严格，标引文献时必须依据一定的规则，并且规则间、规则内各字段间互相联系、互相制约，共同实现标引的功能。句法上的不一致决定了两者转换的艰难。

再次，核心元素集除了其内在性与句法独立性限制了其与 MARC 的转换外，它的可选择性、可扩展性、可修改性、可重复性又促进了转换。但是我们还应看到 DC 元素集的出发点：简单、便于大众使用。从这个原则出发，任意扩展与修改 DC 元素的行为与努力应该被制止。

第四，从规范化的角度来看，MARC 具有完备的理论基础，而核心元素集很难达到 MARC 理论所要求的程度，如主款目、附加款目。MARC 具有一系列的规范化工具，它们只有经过培训的专业人员才能使用；而元数据一般面向的是不具备这些专业的普通大众，也不强求大家使用这些专门化的工具、术语等。

此外，核心集与 MARC 都会因需要而变化，在实际应用中元素与修饰词的数量与使用规则均不尽相同，MARC 本身也有不同的地区版，所以二者之间的对应关系建立起来后并不是一劳永逸的，实际应用中不同的 MARC 地区版和核心集的地区或针对资源特性的扩展集都会有不同的自定义字段和元素需要对应。本文列出的这一对应表是采用标准通用的 MARC 格式的版本以及核心元素集，所以基本上不包含各地自定义的特殊字段与子字段、元素与修饰词。因而，可以说它是可以通用的，各地可以根据通用的映射表加以扩充，形成本地适用的元数据到 MARC 的映射表。

## 参考文献

- [1] Cooperative online resource catalog.  
<http://www.oclc.org/oclc/promo/10520corc/index.htm> (检索时间: 2003-10-20)
- [2] Dillon, Martin. Metadata for web resource: how metadata works on the web.  
[http://lcweb.loc.gov/catdir/bibcontrol/dillon\\_paper.html](http://lcweb.loc.gov/catdir/bibcontrol/dillon_paper.html) (检索时间: 2003-10-20)
- [3] Godby, Jean. A metalanguage for describing Internet resources using the Dublin Core Element Set.  
[http://www.isoc.org/isoc/whatis/conferences/inet/96/proceedings/a2\\_2.htm](http://www.isoc.org/isoc/whatis/conferences/inet/96/proceedings/a2_2.htm) (检索时间: 2003-10-20)
- [4] Hakala, Juha. Dublin Core/FINMARC/GILS crosswalk. Last updated: 01/15/98 (version 1.3).  
<http://www.lib.helsinki.fi/meta/dcficross.htm> (检索时间: 2003-10-20)
- [5] Hakala, Juha. Internet metadata and library cataloging.  
<http://www.lnb.ot/events/ifla/hakala.html> (检索时间: 2003-10-20)
- [6] Hakala, Juha. The Nordic metadata project: final report. June 1998.  
<http://www.lib.helsinki.fi/meta/nmfinal.doc> (检索时间: 2003-10-20)
- [7] Hickey, Thomas. Cooperative online resource catalog explores uses for catalog of Internet resources.  
[http://www.oclc.org/oclc/new/n235/cooperative\\_online\\_resource\\_catalog.htm](http://www.oclc.org/oclc/new/n235/cooperative_online_resource_catalog.htm) (检索时间: 2003-10-20)
- [8] Huthwaite, Ann. AACR2 and its place in the digital world: near-term solutions and long-term direction.  
[http://lcweb.loc.gov/catdir/bibcontrol/huthwaite\\_paper.html](http://lcweb.loc.gov/catdir/bibcontrol/huthwaite_paper.html) (检索时间: 2003-10-20)
- [9] Library of Congress. Metadata, Dublin Core and MARC21: Discussion paper no. 99.  
<http://lcweb.loc.gov/marc/marbi/dp/dp99.html> (检索时间: 2003-10-20)
- [10] Network Development and MARC Standards Office, Library of Congress. Dublin Core/MARC/GILS crosswalk. November, 1999.  
<http://lcweb.loc.gov/marc/dccross.html> (检索时间: 2003-10-20)
- [11] Olson, Nancy B., ed. Cataloging Internet Resource – a manual and practical guide. Second Edition.  
<http://www.oclc.org/oclc/man/9256cat/toc.htm> (检索时间: 2003-10-20)
- [12] Dublin Core Qualifiers.  
<http://purl.org/DC/documents/dcmes-qualifiers/> (检索时间: 2003-10-20)
- [13] Weibel, Stuart. Metadata: the foundations of resources description. D-Lib Magazine, July 1995. (检索时间: 2003-10-20)
- [14] Xu, Amanda. Accessing information on the Internet: feasibility study of MARC21 format and AACR2.  
<http://www.oclc.org/oclc/man/colloq/xu.htm> (检索时间: 2003-10-20)
- [15] Yarra Plenty Regional Library. Cataloging of Internet resources.  
<http://www.yprl.vic.gov.au/yprl/catintres.html> (检索时间: 2003-10-20)

- [16]Zeng, Marcia Lei and Xiao, Long. Mapping metadata elements of different formats. E-libraries Conference 2001 Annual Conference.
- [17]黄纯元. 图书馆与网络信息资源. 中国图书馆学报, 1997(6), p13-19
- [18]黄伟红, 张福炎. 基于 XML/RDF 的 MARC 元数据描述技术. 情报学报, 2000(4), p326-332
- [19]刘嘉. 元数据研究. 博士论文, 北京大学信息管理系, 2000
- [20]刘英梅, 刘赛红. 都柏林核心元数据及其应用. 情报科学, 2000(6), p572-574
- [21]沈芸芸. Metadata 与网络信息资源的控制. 中国图书馆学报, 2000(6), p56-58
- [22]吴建中主编. DC 元数据. 上海: 上海科学技术文献出版社, 2000
- [23]谢琴芳主编. CALIS 联机合作编目手册. 北京: 北京大学出版社, 2000
- [24]张敏, 张晓林. 元数据(Metadata)的发展和相关格式. 四川图书馆学报, 2000,(2),p63-70
- [25]赵华, 车文贤. 网络环境下的信息资源组织. 情报杂志, 2000(1), p58-60