



# 中华人民共和国国家标准

GB/T 34832—2017

---

## 信息与文献 CNMARC 的 XML 表示

Information and documentation—ChinaMarcXchange

(ISO 25577: 2013, Information and documentation—MarcXchange, MOD)

2017-11-01 发布

2018-05-01 实施

---

中华人民共和国国家质量监督检验检疫总局  
中国国家标准化管理委员会 发布

## 目 次

前言 .....	III
引言 .....	IV
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 CNMARC 记录的 XML 模式转换 .....	3
附录 A (资料性附录) CNMARCXML 的一般模式 .....	6
附录 B (资料性附录) 实例 .....	10
参考文献 .....	16

## 前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准使用重新起草法修改采用 ISO 25577:2013《信息与文献 MarcXchange》。

本标准与 ISO 25577:2013 的技术差异及其原因如下：

——关于规范性引用文件，本标准做了具有技术性差异的调整，以适应我国的技术条件，调整的情况集中反映在第 2 章“规范性引用文件”中，具体调整如下：

- 用等同采用国际标准的 GB/T 2901 代替了 ISO 2709；
- 用等同采用国际标准的 GB/T 1988 代替了 ISO/IEC 646；
- 用等同采用国际标准的 GB/T 13000 代替了 ISO/IEC 10646；
- 新增国家标准 GB/T 33286—2016《中国机读书目格式》。

——本标准在第 3 章“术语和定义”中增加了术语“中国机读目录”，即 CNMARC。

——本标准用“CNMARCXML”替代了 ISO 25577:2013 中的“MARCXML”。

——ISO 25577:2013 在 4.1“概述”中规定“子字段标识可采用 ISO 10646 的 8 位编码字符集 BMP row 00(基本拉丁和拉丁-1 补充字符集)”，本标准根据中国机读目录格式 CNMARC 规定改为“子字段代码有且只有一位，可采用大小写字母或数字”。

——本标准为了适应我国 CNMARC 格式的相关规定，将 ISO 25577:2013 4.2“XML 框架结构”中的相关规定进行了调整，具体调整如下：

- 将 4.2 标题改为“CNMARCXML 模式的一般结构”。
- 将 ISO 25577:2013 的 4.2“XML 框架结构”b) 中内容“format(记录格式,可选),标识 MARC 记录的执行格式(如 MARC21、UNIMARC、danMARC2、lbermarc)”改为“format(格式,可选),标识 MARC 记录的执行格式,指定值为‘CNMARC’”。CNMARC 即指中国机读目录格式。
- 将 ISO 25577:2013 的 4.2“XML 框架结构”d) 中内容“controlfield(控制字段):对应 ISO 2709 记录的控制号字段(001)及参考字段(002~009,00A~00Z)”改为“对应 CNMARC 记录的控制号字段(001)及参考字段(002~009,00a~00z,00A~00Z)”,并增加了限制说明语“它仅适用于没有指示符和子字段的 CNMARC 字段”。
- 将 ISO 25577:2013 的 4.2“XML 框架结构”e) 中“datafield”数据字段范围由“001~999 及 00A~ZZZ”修改为“010~999,01a~zzz 及 01A~ZZZ”;将指标符属性由“ind1 to ind9(指示符 1 到指示符 9,可选)”修改为“ind1 和 ind2(指示符 1 和指示符 2,必备)”,同时将对应的图 2 “CNMARCXML 一般结构”中“datafield”的属性修改为“id, ind1, ind2, tag”。

——本标准删除了 ISO 25577:2013 附录 C(资料性)“MarcXchange 结构的应用”。

——本标准将 ISO 25577:2013 附录中样例均修改为符合“CNMARC”格式的资料。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息与文献标准化技术委员会(SAC/TC 4)提出并归口。

本标准起草单位:中国国家图书馆、清华大学图书馆。

本标准主要起草人:高红、汪东波、王洋、延卫平、杨慧。

## 引 言

中国机读目录格式(CNMARC)是我国图书情报界用来记录书目文献元数据的统一标准,它提供了一套完整、详尽、严密的数据表示规范。书目数据是图书馆提供给用户最大的信息资源,但是,由于GB/T 2901 格式的 CNMARC 元数据的专用性,使用者必须依靠专用的客户端并且在特定的软件系统所提供的检索工具中,才能够进行信息的检索。

XML(Extensible Markup Language)是由 W3C(World Wide Web Consortium,万维网联盟)推广、采用的一个新标准,它是 SGML 的一个优化子集。XML 是一种元标记语言,具有良好的扩展性,使用者可按需要扩展、定义新的标记。将 CNMARC 书目记录从 GB/T 2901 格式转换为 XML 格式,可使书目数据成为机器可读的 XML 结构化数据,实现书目数据和因特网上非书目数据的集成,从而使现有的、大量的 CNMARC 数据能够方便地在传统图书馆和数字图书馆中应用,提供面向万维网的有效信息。

早在 2005 年,ISO TC 46/SC 4(信息和文献工作标准化技术委员会/互操作分委员会)曾发布标准草案“ISO/DIS 25577: Information and documentation—MarcXchange”,并推出了一个称为“MarcXchange”的 MARCXML 模式。该模式是一个通用的、能够满足世界范围所有遵循 ISO 2709 的 MARC 格式的通用 XML 模式,它可在 ISO 2709 格式的 MARC 记录和以 XML 编码的 MARC 记录之间进行无损性互转,也可确保 MARC 数据与其他元数据(如 DC 元数据)的兼容。2008 年 12 月 1 日 ISO TC 46/SC 4 正式发布了“ISO 25577: Information and documentation—MarcXchange”第 1 版。2013 年 12 月 15 日正式发布第 2 版,在第 2 版主要增加了嵌套数据元素。

本标准根据 CNMARC 的具体规定,在 ISO 25577:2013 基础上进行了修改和限制,并将 CNMARC 的 XML 模式命名为“CNMARCXML”。采用与 GB/T 2901 格式一致的定义和术语描述,给出一个适用于采用 GB/T 2901 句法但不改变其结构的 CNMARC 记录的通用 XML 模式。

CNMARCXML 结构的核心是一个包含有 CNMARC 数据的简单 XML 编码方案,可用于 MARC 记录的交换,或与其他元数据进行对应转换以及格式校验等,但仍保留 MARC 的语义。在 XML 结构中,将按照如下方式处理 GB/T 2901 的各种基本元素:

- 记录头标以文本字符串表示;
- XML 结构中无目次区,当从 CNMARCXML 转换为 GB/T 2901 格式时,需要重新设置目次区;
- 记录的标识号字段和控制字段以元素表示,字段标识符作为属性;
- 数据字段以元素表示,字段标识符和指示符分别作为其属性;
- 子字段以子元素表示,其子字段代码作为其属性。

# 信息与文献 CNMARC 的 XML 表示

## 1 范围

本标准规定了 CNMARC 记录的 XML 格式,作为网络环境下对 GB/T 2901 格式的一种补充。

本标准阐明了通用的、基于 XML 交换格式的书目记录和其他类型元数据的功能需求,但不定义单个记录的长度或内容,也不为字段标识、指示符、标识符以及执行格式的功能说明赋予任何定义。

本标准适用于使用 CNMARC 格式编制的各种类型记录,包括书目记录、规范记录、分类记录和馆藏记录。

本标准具有以下应用前景:

- XML 环境下完整的 CNMARC 记录表示;
- 用 XML 语法对原始资源进行描述;
- 可作为 METS(元数据编码和传输标准)中描述元数据的扩展编码方案;
- 用于 CNMARC 记录的 XML 交换格式;
- 在 Web 服务中通过 SRU(URL 的查询/检索)等协议传递 CNMARC 记录;
- 用于出版者传输数据;
- 作为各种数据转换或处理的中间格式,如转换、发布、编辑、校验;
- XML 格式的元数据可与对象数据一起封装。

## 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 1988 信息技术 信息交换用七位编码字符集(GB/T 1988—1998,ISO/IEC 646:1991, IDT)

GB/T 2901 信息与文献 信息交换格式(GB/T 2901—2012,ISO 2709:2008, IDT)

GB/T 13000 信息技术 通用多八位编码字符集(UCS)(GB/T 13000—2010,ISO/IEC 10646:2003, IDT)

GB/T 33286—2016 中国机读书目格式

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

#### 机读目录 MARC

用于书目记录及相关信息机读形式存储及交换的系列格式标准。

注: MARC 是机读目录英文全称 MACHine-Readable Cataloguing 的首字母缩写。所有的 MARC 标准均遵循 GB/T 2901。