



中华人民共和国国家标准

GB/T 20092.3—2022

部分代替 GB/T 20092—2013

中文新闻信息内容 第 3 部分：数据交换的 XML 格式

Chinese news information content—
Part 3: XML format for data exchange

2022-10-12 发布

2022-10-12 实施

国家市场监督管理总局
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 缩略语	1
5 规则和约定	1
5.1 XML Schema 格式定义	1
5.2 拼写规则	2
5.3 元素表达及描述要求	2
5.4 XML 组件的命名	2
5.5 XML Schema 的版本和变更约定	2
5.6 编码方式	3
6 公共结构和基本数据类型	3
6.1 目标命名空间声明	3
6.2 基本简单类型	3
6.3 公共结构类型	5
6.4 复杂元数据项类型	16
6.5 元数据相关的 Schema 定义	58
7 关系机制	63
7.1 关系概述	63
7.2 关系定义	64
8 受控词描述机制	69
8.1 概述和分析	69
8.2 受控词表的建模	70
8.3 受控词引用的建模	81
8.4 受控词表的例子	83
9 数字签名	86
9.1 概述	86
9.2 元数据描述	86
9.3 数字签名样例	88
10 文档层	93
10.1 概述	93
10.2 CNML 元素(<CNML>)	94
10.3 词表编目结构	95

10.4	信封结构	97
10.5	稿件容器	129
10.6	稿件关系容器	143
11	稿件层	146
11.1	稿件层定义	146
11.2	稿件层结构	147
11.3	稿件层元数据组	153
12	内容项层	252
12.1	内容项层结构	252
12.2	内容项层实体内容	255
12.3	内容项层元数据组	267
13	内联智能标记	331
13.1	概述	331
13.2	内联智能标记的命名空间	331
13.3	规范说明	331
13.4	内联智能标记类型(cim: IntelligentMarkupType)	331
附录 A (规范性)	XML Schema	341
附录 B (资料性)	应用实例	419
附录 C (资料性)	类继承总图	445
附录 D (规范性)	内联智能标记 XML Schema	446
参考文献	449

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

本文件是 GB/T 20092《中文新闻信息内容》的第 3 部分。GB/T 20092 已经发布了以下部分：

- 第 1 部分：概念模型；
- 第 2 部分：新闻元数据；
- 第 3 部分：数据交换的 XML 格式。

本文件代替 GB/T 20092—2013《中文新闻信息置标语言》中的第 5 章“规则和约定”、第 7 章“CNML 公共结构和基本数据类型”、第 8 章“CNML 的关系机制”、第 9 章“受控词描述机制”、第 10 章“数字签名”、第 11 章“CNML 文档层结构”、第 12 章“CNML 稿件层”、第 13 章“CNML 内容项层”、第 14 章“CNML 内联智能标记”，附录 A（资料性附录）“CNML 应用实例”、附录 B（资料性附录）“CNML 标准元素中英文对照表”、附录 C（资料性附录）“CNML 类继承总图”、附录 E（规范性附录）“CNML XML Schema”、附录 F（规范性附录）“内联智能标记 XML Schema”，与 GB/T 20092—2013 相比，除结构调整和编辑性改动外，主要技术变化如下：

- a) 增加了“XML Schema 格式定义”（见 5.1）和“编码方式”（见 5.6）；
- b) 删除了有关“CNML”的词语（见 2013 年版的第 6 章、第 10 章、第 11 章、第 12 章、第 13 章等）；
- c) 更改了“视音频内容项元数据组类型”的部分内容（见 12.3.4.5, 2013 年版的 13.3.4.5、13.4.4.6）；
- d) 删除了“CNML 标准元素中英文对照表”（见 2013 年版的附录 B）。

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专利的责任。

本文件由全国中文新闻信息标准化技术委员会（SAC/TC 352）提出并归口。

本文件起草单位：新华通讯社、清华大学、国家应用软件产品质量检验检测中心、北京中联合超高清协同技术中心有限公司、北京北大方正电子有限公司、中国科学院计算机网络信息中心、北京智途云天科技有限公司、视觉（中国）文化发展股份有限公司、北京北科光大信息技术股份有限公司、中国人民大学、中国传媒大学、中国大百科全书出版社有限公司、杭州前方信息技术有限公司、东南大学、新华网股份有限公司、中国互联网新闻中心、南方报业传媒集团、重庆日报报业集团、青苹果数据中心有限公司、北京中科大洋科技发展股份有限公司。

本文件主要起草人：武国卫、黄菁、张鹏、王威、王付生、朱建波、朱艳华、丁峰、王骏清、王堃越、许向东、陈锐、张弛、曹学军、刘万福、杨鹏、杨溟、王建平、何崑、张晓伟、张晴、付蓉、路海燕、张贤良。

本文件所代替文件的历次版本发布情况为：

- 2006 年首次发布为 GB/T 20092—2006，2013 年第一次修订；
- 本次为第二次修订。

引 言

GB/T 20092—2013 自发布实施以来,已在很多新闻媒体单位的应用系统中使用。近年来,在标准实施过程中不断涌现一些新的应用场景,产生一些新的应用需求,新闻媒体的部分应用系统也逐渐不再使用 XML 格式,使用 JSON 格式传输稿件的应用越来越多。因为 GB/T 20092—2013 定义的技术内容主要是基于 XML 格式定义的,已经逐渐不适应新闻媒体的应用需求,所以本次修订的主要目的是避免将标准的内容局限于某一种技术,把中文新闻信息内容的概念模式和新闻元数据独立出来,并将新闻信息内容的 XML 格式和 JSON 格式单独进行规范和定义,从而使标准的范围更加明确,定义更加清晰。

GB/T 20092 旨在定义普遍适用于中文新闻信息内容采集、编辑、生成、发布、交换、存储、检索、评估反馈等新闻信息生命周期的各个环节的描述规范,为适应最新技术在新闻媒体中的实际需要,使标准更具有可用性,将标准名称修改为《中文新闻信息内容》,拟由四个部分构成。

- 第 1 部分:概念模型。目的在于建立新闻行业中文新闻信息的抽象概念模型,规范中文新闻信息文档的总体结构和描述方式,为新闻内容使用者提供指导。
- 第 2 部分:新闻元数据。目的在于规范描述中文新闻信息内容的各类元数据,便于在新闻媒体行业信息系统中实现信息内容的统一表达,一致描述。
- 第 3 部分:数据交换的 XML 格式。目的在于规定基于 XML 的中文新闻信息数据交换格式,为新闻媒体信息系统的建设提供技术支持。
- 第 4 部分:数据交换的 JSON 格式。目的在于规定基于 JSON 的中文新闻信息数据交换格式,为新闻媒体信息系统的建设提供技术支持。

中文新闻信息内容

第 3 部分:数据交换的 XML 格式

1 范围

本文件规定了中文新闻信息数据交换的 XML 格式,包括规则和约定、公共结构和基本数据类型、关系机制、受控词描述机制、数字签名、文档层结构、稿件层结构、内容项层结构和内联智能标记。

本文件适用于新闻信息内容的“采集、编辑、生成、发布、交换、存储、检索、评估反馈”等新闻信息生命周期的各个环节。也适用于包括报刊、广播、电视、通讯社、新闻网站、社交媒体等新闻内容提供商及媒体应用与研究机构。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;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 4880.2 语种名称代码 第 2 部分:3 字母代码

GB/T 4881 中国语种代码

GB/T 7408 数据元和交换格式 信息交换 日期和时间表示法

GB/T 20092.1—2022 中文新闻信息内容 第 1 部分:概念模型

3 术语和定义

GB/T 20092.1—2022 界定的术语和定义适用于本文件。

4 缩略语

下列缩略语适用于本文件。

URI:统一资源标识符(Uniform Resource Identifier)

URL:统一资源定位器(Uniform Resource Locator)

URN:统一资源名称(Uniform Resource Name)

UTF-8:8 位统一码转换格式(8-bit Unicode Transformation Format)

XML:可扩展置标语言(Extensible Markup Language)

5 规则和约定

5.1 XML Schema 格式定义

本文件按照 GB/T 20092.1—2022 中 4.1 的定义,给出了中文新闻信息数据交换的 XML 格式描