

WS

中华人民共和国卫生行业标准

WS/T 305—2023
代替 WS/T 305-2009

卫生健康信息数据集元数据标准

Metadata specification of health information dataset

2023-08-07 发布

2024-02-01 实施

中华人民共和国国家卫生健康委员会 发布

前 言

本标准为你推荐性标准。

本标准代替WS/T 305—2009 《卫生信息数据集元数据规范》。与WS/T 305—2009相比，主要为编辑性修改。

本标准由国家卫生健康标准委员会卫生健康信息标准专业委员会负责技术审查和技术咨询，由国家卫生健康委统计信息中心负责协调性和格式审查，由国家卫生健康委规划发展与信息化司负责业务管理、法规司负责统筹管理。

本标准起草单位：中国人民解放军总医院、国家卫生健康委统计信息中心、中国人民解放军空军军医大学。

本标准主要起草人：刘丽华、王珊、张黎黎、李林、王霞、刘建超、胡建平。

卫生健康信息数据集元数据标准

1 范围

本标准规定了数据集元数据内容框架、卫生健康信息数据集核心元数据、卫生健康信息数据集参考元数据、引用信息与代码表。

本标准适用于作为卫生健康信息数据集属性的统一规范化描述，也可用于卫生健康领域针对数据集制定专用元数据标准的依据。

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标准必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本标准；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

GB/T 2260 中华人民共和国行政区划代码

GB/T 7408 数据元和交换格式 信息交换 日期和时间表示方法

WS/T 306 卫生健康信息数据集分类与编码规则

RFC 2396 统一资源定义符：通用语法

3 术语和缩略语

下列术语和缩略语适用于本标准。

3.1 术语

3.1.1

元数据 metadata

定义和描述其他数据的数据。

3.1.2

数据集 dataset

具有一定主题，可以标识并可以被计算机化处理的数据集合。

3.1.3

元数据元素 metadata element

元数据的基本单元。

3.1.4

元数据实体 metadata entity

一组说明数据相同特性的元数据元素。可以包含一个或若干个元数据实体。

3.1.5

元数据子集 metadata section

元数据的子集合，由相关的元数据子集和元数据元素组成。

3.1.6

数据元 data element

用一组属性规定其定义、标识、表示和允许值的数据单元。

3.2 缩略语

XML: 可扩展标记语言 (Extensible Markup Language)

4 数据集概述

4.1 数据集基本概念

对数据集概念的详细解释可见WS/T 306。

4.2 卫生健康信息数据集

对卫生健康信息数据集的阐述可见 WS/T 306。

5 卫生健康信息数据集元数据概述

5.1 元数据内容组成

卫生健康信息数据集元数据规定了两个层次的元数据内容的集合：核心元数据与参考元数据，其中参考元数据包含了核心元数据，它们之间的内容包含关系见图 1。

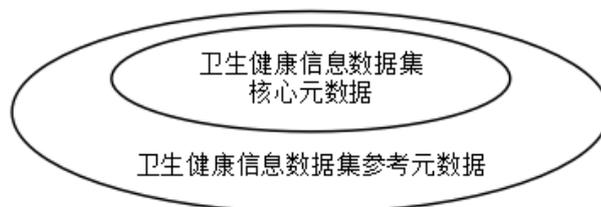


图 1 核心元数据与参考元数据的内容包含关系

5.2 元数据结层次构

元数据结构：元数据元素、元数据实体和元数据子集。

元数据元素是元数据的最基本的信息单元。例如，数据集名称、数据集标识符、元数据创建日期等，是最基本的属性信息单元，用元数据元素来表示。

元数据实体是同类元数据元素的集合，用于一些需要组合若干个更加基本的信息来表达的属性。例如“数据集提交和发布方”需要“单位名称”、“联系人”、“联系电话”、“通讯地址”等若干个基本信息来说明，而数据集“关键词说明”需要“关键词”和“词典名称”来说明，对于“数据集提交和发布方”和“关键词说明”这类属性用元数据实体来表示。

元数据子集由共同说明数据集某一类属性的元数据元素与元数据实体组成，例如标识信息、内容信息、分发信息等。

5.3 元数据的摘要描述

5.3.1 元数据的摘要描述符

主要采用摘要描述方法定义和描述卫生健康信息数据集的元数据内容。摘要描述的方法通过表1所列的描述符针对元数据的元素、实体与子集进行定义。“约束”表示一个描述符是“必选” (M)，还是“条件选” (C)，或者是“可选” (O)，摘要描述符见表1。

表 1 摘要描述符

描述符	约束	备注
中文名称	M	
定义	M	
英文名称	M	
数据类型	M	
取值规则	C	描述元数据元素时必选，描述元数据子集、实体时不选
短名	M	
子元素	C	描述元数据子集、实体时必选，描述元数据元素时不选
注解	M	

5.3.2 中文名称

元数据子集、实体或元素的中文名称，即为核心元数据的元数据实体或元数据元素的中文名称。

5.3.3 定义

描述元数据子集、实体或元素的基本内容，是卫生健康信息数据集某个属性和特征的概念与说明。

5.3.4 英文名称

元数据子集、实体或元数据元素的英文名称，一般使用英文全称。

英文名称中，词汇均采用无缝连写。元数据子集与实体英文名称的每个词汇的首字母均为大写，元数据元素英文名称的首词汇全部小写，其余词汇的首字母采用大写。

5.3.5 数据类型

对元数据的有效值域和允许对该值域内的值进行有效操作的规定。

元数据实体与元数据子集的数据类型为复合型，元数据元素的数据类型包括为数值型、布尔型、字符串、日期型和复合型等。

当元数据元素的数据类型为复合型的时候，表明该元数据元素使用了引用信息中的引用数据类型。

引用数据类型是由一组元素组成、能被重复引用的元素集合。例如，“数据集提交和发布方”和“元数据维护方”两个属性，要共同用到“单位名称”、“联系人”、“联系电话”、“通讯地址”、“邮政编码”和“电子邮件地址”这一组子元素来表示，为了表达的方便，卫生健康信息元数据定义了一个叫做“负责方”的数据类型见本标准 9.1.1 条款，供“数据集提交和发布方”和“元数据维护方”进行引用。

5.3.6 取值规则

元数据子集、实体或元素取值方式的说明。

5.3.7 短名

元数据元素、实体或子集的英文缩写名称，具体缩写规则如下：

- a) 短名在本标准范围内必须唯一。
- b) 对存在国际或行业惯用英文缩写词汇的元数据元素或元数据实体，采用惯用英文缩写作短名。
- c) 对于根据英文名称或自定义的短名，在保证其（本标准范围中的所有短名范围内）唯一性的情况下取每个词汇的前 3 位字母作为其短名，当如此取词时发现不能够保证短名标识的唯一性时，应当延展取词的位数，通常增加 1 位。如果仍然不能满足唯一性要求，则照此顺延。
- d) 元数据子集、元素和元数据实体的短名中，词汇均采用无缝连写。元数据子集与实体短名

中每个词汇的首字母均为大写；元数据元素短名的首词汇全部小写，其余词汇的首字母采用大写。

5.3.8 子元素

子元素是通过一定的表示规则以确定一个元数据实体中包含的下一级的元数据实体或元数据元素。表示规则为：“标识符 = 表达式”。表达式中各符号的含义见表 2。

表 2 表达式的符号含义

符号	含 义
=	由……替换、生成，由……组成
+	与
	或（选择）——在由“ ”分开的两项之中选择其一
0{a}1	表示 {} 中的元数据元素 a 为可选/条件必选项，且最大出现次数为 1；若为条件必选项，约束/条件参见其注解
0{a}n	表示 {} 中的元数据元素 a 为可选/条件必选项，且最大出现次数为 N；若为条件必选项，约束/条件参见其注解
a	表示元数据元素 a 为必选项，且最大出现次数为 1
1{a}n	表示 {} 中的元数据元素 a 为必选项，且最大出现次数为 N

在子元素表示中，{} 中均使用元数据元素或实体的中文名称。

例如：

子元素：关键词说明 =
 $1\{\text{关键词}\}n +$
 $0\{\text{词典名称}\}1$

5.3.9 注解

对元数据子集、实体或元素的含义的进一步解释，包括该元数据的约束/条件（必选、可选或条件必选）和最大出现次数。当该元数据为条件必选时，应注明其约束条件。

5.3.9.1 约束/条件

以“必选项”、“可选项”、“条件可选”等文字说明元数据子集、实体或元素是“必须 (M)”、“可选 (O)”或者“条件可选 (C)”。

可选元数据子集或元数据实体可以有必选元数据实体或元素。但这些必选的实体或元素只有当上一级的可选子集或实体被选用时才成为必选的；如果上一级的可选元数据子集或实体未被选用，则其所包含的必选实体或元素也不选用。

5.3.9.2 最大出现次数

说明元数据子集、实体或元素可以重复出现的次数。只允许出现一次的用“1”表示，允许重复出现（但重复的次数不是固定值）的用“N”表示。不为 1 的固定出现次数用相应的数字表示，如“2”、“3”、“4”等。

5.4 适用功能

5.4.1 数据集元数据著录

对具体的卫生健康信息数据集选择元数据内容进行描述时，以核心元数据作为著录（根据必选、可选等约束条件）的基础，然后根据需要可以选择参考元数据的内容进行著录，见图 2：

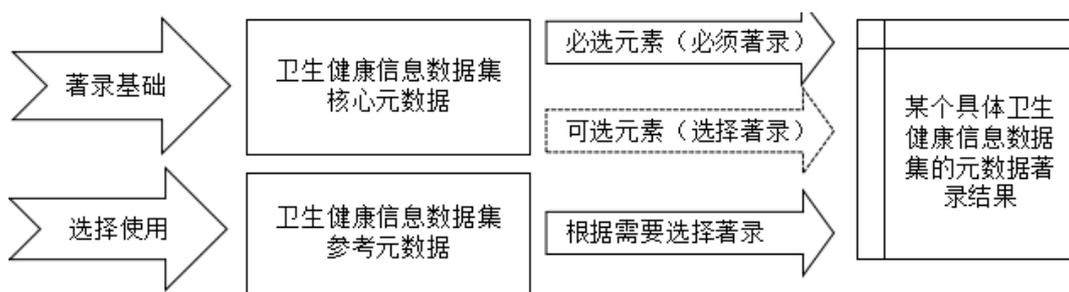


图2 直接作为著录标准的使用方式

用户针对卫生健康信息数据集进行元数据著录的示例，可参照本标准附录A（以只著录核心元数据内容为例）。

5.4.2 制定专用标准的依据

在卫生健康领域内，以数据集为对象制定专用的元数据标准时，可依照本标准进行制定。

所制定的数据集专用元数据标准，必须包含核心元数据的内容，同时将参考元数据的内容作为扩展内容的参考，使用方式见图3：

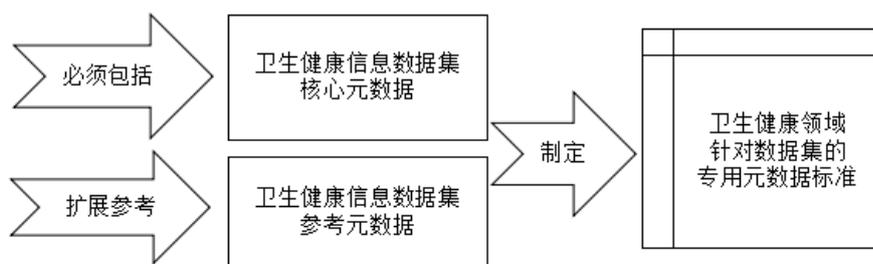


图3 作为制定专用标准依据时的使用方式

要在卫生健康信息参考元数据的基础上继续进行元数据内容的扩展时，应当符合附录B的要求。

6 元数据内容框架

卫生健康信息数据集元数据的全集，即参考元数据的内容规定了7个元数据子集，分别如下：

- a) 元数据标识信息（子集）：包含元数据本身的基本信息，如名称、维护方、创建日期等。
- b) 数据集标识信息（子集）：包含元数据所描述的数据集的基本信息，包括名称、标识符、提交或发布方等。
- c) 内容信息（子集）：内容信息提供数据内容特征的描述信息，包括摘要、特征数据元等。
- d) 分发信息（子集）：分发信息包含有关资源分发者的信息以及用户获取资源的途径。
- e) 数据质量信息（子集）：数据生产者有关数据集数据志信息的一般说明。
- f) 限制信息（子集）：包含对数据集和元数据进行访问和使用的限制信息。
- g) 维护信息（子集）：包含关于元数据或数据集的更新频率及更新范围的信息。

参照UML类图中聚集关系与依赖关系的表示方法，概要描述了卫生健康信息数据集参考元数据的内容框架（见图4）。

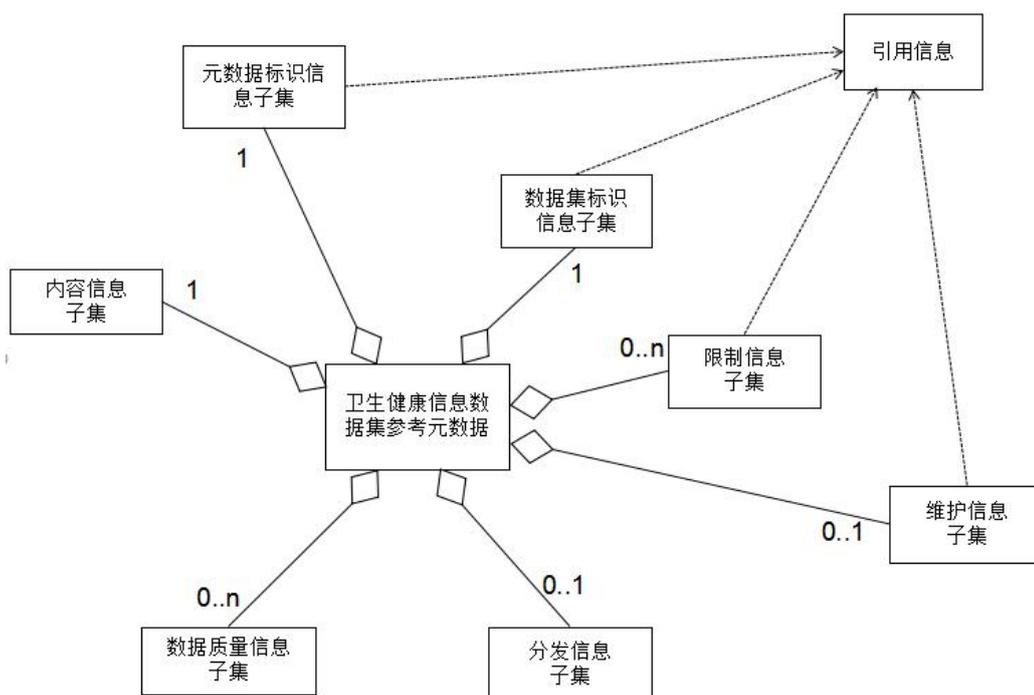


图 4 卫生健康信息参考元数据内容构成

7个子集是参考元数据这一整体的组成部分，在图4中标记出了各自的可出现次数。元数据标识信息（子集）、数据集标识信息（子集）、限制信息（子集）和维护信息（子集）中有元数据元素使用到了引用信息中由本标准定义的数据类型，所以它们对引用信息存在依赖关系。

7 卫生健康信息数据集核心元数据

7.1 概述

卫生健康信息核心元数据包括8个必选元数据元素或实体，10个可选元数据元素与实体。卫生健康信息数据集核心元数据内容见表3，其中的第四列“约束”是指该元数据元素或元数据实体是“必选（M）”，或者是“可选（O）”，第五列是说明核心元数据在参考元数据中的相应位置。

表 3 卫生健康信息数据集核心元数据内容

序号	核心元数据内容	元素与实体	约束	在参考元数据中的位置
1	数据集名称	元素	M	数据集标识信息子集
2	数据集标识符	元素	M	数据集标识信息子集
3	数据集摘要	元素	M	内容信息子集
4	数据集提交或发布方	元素	M	数据集标识信息子集
5	关键词说明	实体	M	数据集标识信息子集
6	数据集语种	元素	M	数据集标识信息子集
7	数据集特征数据元	元素	O	内容信息子集
8	数据集发布日期	元素	O	发布信息子集
9	数据集发布格式	实体	O	发布信息子集
10	在线访问地址	元素	O	发布信息子集
11	数据集分类	实体	O	数据集标识信息子集
12	相关环境说明	元素	O	数据集标识信息子集
13	元数据创建日期	元素	M	元数据标识信息子集
14	元数据标识符	元素	M	元数据标识信息子集
15	元数据负责人	元素	O	元数据标识信息子集
16	元数据标准名称	元素	O	元数据标识信息子集
17	元数据标准版本	元素	O	元数据标识信息子集
18	元数据更新日期	元素	O	元数据标识信息子集

7.2 摘要描述

7.2.1 数据集名称（必选）

定 义：能够简要描述卫生健康信息数据集主题与内容的标题

英文名称：datasetTitle

数据类型：字符串

取值规则：自由文本

短 名：dataTitle

注 解：必选项；最大出现次数为 1

7.2.2 数据集标识符（必选）

定 义：卫生健康信息数据集的唯一标识符

英文名称：datasetIdentifier

数据类型：字符串

取值规则：按照“提交机构-编号-版本号”的命名规则进行赋值。提交机构使用机构标准名称，版本号用一位整数加两位小数的表示方式，例如“2.10”（版本）。

短 名：dataID

注 解：必选项；最大出现次数为 1

7.2.3 数据集摘要（必选）

定 义：数据集内容的简单说明

英文名称：abstract

数据类型：字符串

取值规则：自由文本

短 名：abstract

注 解：必选项；最大出现次数为 1

7.2.4 数据集提交或发布方（必选）

定 义：提交或发布数据集，并对数据集的真实性、正确性负责的单位或部门

英文名称：datasetContact

数据类型：复合型

取值规则：负责方<<数据类型>>（见本标准 9.1.1）

短 名：dataContact

注 解：必选项；最大出现次数为 N

7.2.5 关键词说明（必选）

定 义：关键词种类、类型和参考资料

英文名称：DescriptiveKeywords

数据类型：复合型

短 名：DescKeys

子 元 素：关键词说明 =
 $1\{\text{关键词}\}n +$
 $0\{\text{词典名称}\}1$

注 解：必选项；最大出现次数为 N

7.2.5.1 关键词

定 义：用于描述数据集主题的通用词、形式化词或短语

英文名称：keyword

数据类型：字符串

取值规则：自由文本
短 名：keyword
注 解：必选项；最大出现次数为 N

7.2.5.2 词典名称

定 义：正式注册的词典名，或类似的权威关键词资料名称
英文名称：thesaurusName
数据类型：字符型
取值规则：自由文本
短 名：thesaName
注 解：可选项；最大出现次数为 1

7.2.6 数据集语种（必选）

定 义：数据集采用的语言
英文名称：datasetLanguage
数据类型：字符串
取值规则：语种代码<<代码表>>（见本标准 9.2.1）
短 名：dataLang
注 解：必选项；最大出现次数为 N

7.2.7 数据集特征数据元（可选）

定 义：卫生健康信息数据集中，能够表达数据集核心内容与特征资源的数据元列举
英文名称：characteristicDataElement
数据类型：字符串
取值规则：自由文本，数据元之间使用英文状态下的逗号分开
短 名：charaDataEle
注 解：可选项；最大出现次数为 1

7.2.8 数据集发布日期（可选）

定 义：卫生健康信息数据集进行提交或发布的日期
英文名称：datasetPublicationDate
数据类型：日期型
取值规则：按照 GB/T 7408-2005 规定，格式应当为 YYYY-MM-DD
短 名：dataPubDate
注 解：可选项；最大出现次数为 1

7.2.9 数据集分发格式（可选）

定 义：分发数据的格式说明
英文名称：DistributionFormat
数据类型：复合型
短 名：DistFormat
子 元 素：分发格式 =
分发格式名称+
0{版本}1
注 解：可选项；最大出现次数为 N

7.2.9.1 分发格式名称

定 义：数据集分发格式名称
英文名称：formatName

数据类型：字符串
 取值规则：自由文本，默认为文件扩展名称
 短 名：formName
 注 解：必选项；最大出现次数为 1

7.2.9.2 版本

定 义：数据集分发格式所对应的软件版本（日期、版本号等）
 英文名称：formatVersion
 数据类型：字符串
 取值规则：自由文本
 短 名：formVer
 注 解：可选项；最大出现次数为 1

7.2.10 在线访问地址（可选）

定 义：可以对数据集进行在线访问或获取的信息
 英文名称：online
 数据类型：字符型
 取值规则：自由文本，按照 RFC 2396 规定填写
 短 名：online
 注 解：可选项；最大出现次数为 N

7.2.11 数据集分类（可选）

定 义：数据集的分类信息
 英文名称：DatasetCategory
 数据类型：复合型
 短 名：DataCat
 子 元 素：数据集分类 =
 类目名称+
 类目编码+
 分类标准
 注 解：可选项；最大出现次数为 N

7.2.11.1 类目名称

定 义：对应于所使用的某种分类方法所得到的具体类目名称
 英文名称：categoryName
 数据类型：字符串
 取值规则：自由文本
 短 名：cateName
 注 解：必选项；最大出现次数为 1

7.2.11.2 类目编码

定 义：类别名称对应的编码
 英文名称：categoryCode
 数据类型：字符串
 取值规则：自由文本
 短 名：cateCode
 注 解：必选项；最大出现次数为 1

7.2.11.3 分类标准

定 义：所依据的分类标准名称
 英文名称：categyStandard
 数据类型：字符串
 取值规则：自由文本
 短 名：cateStd
 注 解：必选项；最大出现次数为 1

7.2.12 相关环境说明（可选）

定 义：说明数据集生产的处理环境，包括软件、计算机操作系统、文件名和数据量等
 英文名称：environmentDescription
 数据类型：字符串
 取值规则：自由文本
 短 名：envirDesc
 注 解：可选项；最大出现次数为 1

7.2.13 元数据创建日期（必选）

定 义：创建、采集元数据的日期
 英文名称：metadataDateStart
 数据类型：日期型
 取值规则：按照 GB/T 7408-2005 规定，格式应当为 YYYY-MM-DD
 短 名：mdDateStart
 注 解：必选项；最大出现次数为 1

7.2.14 元数据标识符（必选）

定 义：元数据的唯一标识
 英文名称：metadataIdentifier
 数据类型：字符串
 取值规则：在 7.2.2 的数据集标识符取值前加“MD-”字样。
 短 名：mdID
 注 解：必选项；最大出现次数为 1；标识符须唯一、由字母[含下划线（_）短划线（-）点（.）斜线（/）逗号（,）和空格（ ）]或数字组成

7.2.15 元数据负责方（可选）

定 义：提交元数据并对元数据负责的部门和单位
 英文名称：metadataContact
 数据类型：复合型
 取值规则：负责方<<数据类型>>（见本标准 9.1.1）
 短 名：mdContact
 注 解：可选项；最大出现次数为 N

7.2.16 元数据标准名称（可选）

定 义：执行的元数据标准名称
 英文名称：metadataStandardName
 数据类型：字符串
 取值规则：自由文本
 短 名：mdStdName
 注 解：可选项；最大出现次数为 1

7.2.17 元数据标准版本（可选）

定 义：执行的元数据标准版本
 英文名称：metadataStandardVersion
 数据类型：字符串
 取值规则：自由文本
 短 名：mdStdVer
 注 解：可选项；最大出现次数为 1

7.2.18 元数据更新日期（可选）

定 义：更新元数据的日期
 英文名称：metadataDateUpdata
 数据类型：日期型
 取值规则：按照 GB/T 7408-2005 规定，格式应当为 YYYY-MM-DD
 短 名：mdDateUpd
 注 解：可选项；最大出现次数为 1

8 卫生健康信息数据集参考元数据

8.1 元数据根实体

卫生健康信息参考元数据由 7 个元数据子集构成，继续采用 5.3 中的规则来对参考元数据的各个子集、实体和元素进行摘要描述。卫生健康信息参考元数据的根实体描述如下：

中文名称：元数据（根实体）
 定 义：卫生健康信息数据集参考元数据的根实体
 英文名称：Metadata
 数据类型：复合型
 短 名：Metadata
 子 元 素：元数据 =

元数据标识信息子集 +
 数据集标识信息子集 +
 内容信息子集 +
 0{分发信息子集}1 +
 0{数据质量信息子集}n +
 0{限制信息子集}n +
 0{维护信息子集}1

8.2 元数据子集

8.2.1 元数据标识信息子集

定 义：关于元数据本身的基本信息
 英文名称：MetadataIdentificationInfo
 数据类型：复合型
 短 名：MdIDInfo
 子 元 素：元数据标识信息子集 =
 元数据创建日期 +
 元数据标识符 +
 0{元数据负责方}n +
 0{元数据更新日期}1 +
 0{元数据标准名称}1 +
 0{元数据标准版本}1 +

注 解：必选项，最大出现次数为 1

8.2.1.1 元数据创建日期

定 义：创建、采集元数据的日期
 英文名称：metadataDateStart
 数据类型：日期型
 取值规则：按照 GB/T 7408-2005 规定，格式应当为 YYYY-MM-DD
 短 名：mdDateSt
 注 解：必选项；最大出现次数为 1

8.2.1.2 元数据标识符

定 义：元数据的唯一标识
 英文名称：metadataIdentifier
 数据类型：字符串
 取值规则：在 7.2.2 的数据集标识符取值前加“MD-”字样。
 短 名：mdID
 注 解：必选项；最大出现次数为 1；标识符须唯一、由字母[含下划线（_）短划线（-）点（.）斜线（/）逗号（,）和空格（ ）]或数字组成

8.2.1.3 元数据负责方

定 义：提交元数据并对元数据负责的部门和单位
 英文名称：metadataContact
 数据类型：复合型
 取值规则：负责方<<数据类型>>（见本标准 9.1.1）
 短 名：mdContact
 注 解：可选项；最大出现次数为 N

8.2.1.4 元数据更新日期

定 义：更新元数据的日期
 英文名称：metadataDateUpdata
 数据类型：日期型
 取值规则：按照 GB/T 7408-2005 规定，格式应当为 YYYY-MM-DD
 短 名：mdDateUpd
 注 解：可选项；最大出现次数为 1

8.2.1.5 元数据标准名称

定 义：执行的元数据标准名称
 英文名称：metadataStandardName
 数据类型：字符串
 取值规则：自由文本
 短 名：mdStdName
 注 解：可选项；最大出现次数为 1

8.2.1.6 元数据标准版本

定 义：执行的元数据标准版本
 英文名称：metadataStandardVersion
 数据类型：字符串
 取值规则：元数据标准版本的规范写法
 短 名：mdStdVer
 注 解：可选项；最大出现次数为 1

8.2.2 数据集标识信息子集

定 义：元数据所描述的数据集的基本信息

英文名称：DatasetIdentificationInfo

数据类型：复合型

短 名：dataIdInfo

子 元 素：数据集标识信息子集 =

- 数据集名称 +
- 数据集标识符+
- 0{目的}1 +
- 0{可信度}n +
- 1{数据集提交或发布方}n +
- 1{关键词说明}n +
- 1{数据集语种}n +
- 1{数据集分类}n +
- 0{相关环境说明}1

注解：必选项，最大出现次数为 1

8.2.2.1 数据集名称

定 义：能够简要描述卫生健康信息数据集主题与内容的标题

英文名称：datasetTitle

数据类型：字符串

取值规则：自由文本

短 名：dataTitle

注 解：必选项；最大出现次数为 1

8.2.2.2 数据集标识符

定 义：卫生健康信息数据集的唯一标识符

英文名称：datasetIdentifier

数据类型：字符串

取值规则：按照“提交机构-编号-版本号”的命名规则进行赋值。提交机构使用机构标准名称，版本号用一位整数加两位小数的表示方式，例如“2.10”（版本）。

短 名：dataID

注 解：必选项；最大出现次数为 1

8.2.2.3 目的

定 义：数据集开发的目的说明

英文名称：purpose

数据类型：字符串

取值规则：自由文本

短 名：Purp

注 解：可选项；最大出现次数为 1

8.2.2.4 可信度

定 义：对数据集中数据内容可靠性的说明

英文名称：credit

数据类型：字符串

取值规则：自由文本

短 名：Credit

注 解：可选项；最大出现次数为 N

8.2.2.5 数据集提交或发布方

定 义：提交或发布数据集，并对数据集的真实性、正确性负责的单位或部门

英文名称：datasetContact

数据类型：复合型

取值规则：负责方<<数据类型>>（见本标准 9.1.1）

短 名：dataContact

注 解：必选项；最大出现次数为 N

8.2.2.6 关键词说明

定 义：关键词种类、类型和参考资料

英文名称：DescriptiveKeywords

数据类型：复合型

短 名：DescKeys

子 元 素：关键词说明 =

1{关键词}n +

0{词典名称}1

注 解：必选项；最大出现次数为 N

8.2.2.6.1 关键词

定 义：用于描述数据集主题的通用词、形式化词或短语

英文名称：keyword

数据类型：字符串

取值规则：自由文本

短 名：keyword

注 解：必选项；最大出现次数为 N

8.2.2.6.2 词典名称

定 义：正式注册的词典名，或类似的权威关键词资料名称

英文名称：thesaurusName

数据类型：字符型

取值规则：自由文本

短 名：thesaName

注 解：可选项；最大出现次数为 1

8.2.2.7 数据集语种

定 义：数据集采用的语言

英文名称：datasetLanguage

数据类型：字符串

取值规则：语种代码<<代码表>>（见本标准 9.2.1）

短 名：dataLang

注 解：必选项；最大出现次数为 N

8.2.2.8 数据集分类

定 义：数据集的分类信息

英文名称：DatasetCategory

数据类型：复合型

短 名：DataCat

子 元 素：数据集分类 =

类目名称+

类目编码+
分类标准

注 解：可选项；最大出现次数为 N

8.2.2.8.1 类目名称

定 义：对应于所使用的某种分类方法所得到的具体类目名称

英文名称：categoryName

数据类型：字符串

取值规则：自由文本

短 名：cateName

注 解：必选项；最大出现次数为 1

8.2.2.8.2 类目编码

定 义：类别名称对应的编码

英文名称：categoryCode

数据类型：字符串

取值规则：自由文本

短 名：cateCode

注 解：必选项；最大出现次数为 1

8.2.2.8.3 分类标准

定 义：所依据的分类标准名称

英文名称：categoryStandard

数据类型：字符串

取值规则：自由文本

短 名：cateStd

注 解：必选项；最大出现次数为 1

8.2.2.9 相关环境说明

定 义：说明数据集生产的处理环境，包括软件、计算机操作系统、文件名和数据量等

英文名称：environmentDescription

数据类型：字符串

取值规则：自由文本

短 名：envirDesc

注 解：可选项；最大出现次数为 1

8.2.3 内容信息子集

定 义：卫生健康信息数据集内容特征的描述信息

英文名称：Content

数据类型：复合型

短 名：Cont

子 元 素：内容信息子集 =
数据集摘要+
0{数据集特征数据元}1 +
0{数据获取手段 }1 +
0{数据描述对象 }n +

注 解：必选项，最大出现次数为 1

8.2.3.1 数据集摘要

定 义：数据集内容的简单说明
 英文名称：abstract
 数据类型：字符串
 取值规则：自由文本
 短 名：abstract
 注 解：必选项；最大出现次数为 1

8.2.3.2 数据集特征数据元

定 义：卫生健康信息数据集中，能够表达数据集核心内容与特征资源的数据元列举。
 英文名称：characteristicDataElement
 数据类型：字符串
 取值规则：自由文本，数据元之间使用英文状态下的逗号分开。
 短 名：charaDataEle
 注 解：可选项；最大出现次数为 1

8.2.3.3 数据获取手段

定 义：数据集产生的手段、途径，包括监测、流调、电子病历等
 英文名称：dataFrom
 数据类型：字符串
 取值规则：自由文本
 短 名：dataFrom
 注 解：可选项；最大出现次数为 1

8.2.3.4 数据描述对象

定 义：数据集所描述的对象，包括环境、食品药品、人群、人体、器官等
 英文名称：dataTarget
 数据类型：字符串
 取值规则：自由文本
 短 名：dataTarget
 注 解：可选项；最大出现次数为 N

8.2.4 分发信息子集

定 义：提供数据集的分发格式和分发方式的信息
 英文名称：DistributionInfo
 数据类型：复合型
 短 名：DistInfo
 子 元 素：分发信息子集 =

0{数据集发布格式}n +
 0{数据集发布日期}n +
 0{在线访问地址}n +
 0{离线发布介质}n

注 解：可选项；最大出现次数为 1

8.2.4.1 数据集分发格式

定 义：分发数据的格式说明
 英文名称：DistributionFormat
 数据类型：复合型
 短 名：DistFormat
 子 元 素：分发格式 =

分发格式名称+
0{版本}1

注 解：可选项；最大出现次数为 N

8.2.4.1.1 分发格式名称

定 义：数据集分发格式名称

英文名称：formatName

数据类型：字符串

取值规则：自由文本，默认为文件扩展名称

短 名：formName

注 解：必选项；最大出现次数为 1

8.2.4.1.2 版本

定 义：数据集分发格式所对应的软件版本（日期、版本号等）

英文名称：formatVersion

数据类型：字符串

取值规则：自由文本

短 名：formVer

注 解：可选项；最大出现次数为 1

8.2.4.2 数据集发布日期

定 义：卫生健康信息数据集进行提交或发布的日期

英文名称：datasetPublicationDate

数据类型：日期型

取值规则：按照 GB/T 7408 规定，格式应当为 YYYY-MM-DD

短 名：dataPubDate

注 解：可选项；最大出现次数为 1

8.2.4.3 在线访问地址

定 义：可以对数据集进行在线访问或获取的信息

英文名称：online

数据类型：字符型

取值规则：自由文本，按照 RFC 2396 规定填写

短 名：online

注 解：可选项；最大出现次数为 N

8.2.4.4 离线发布介质

定 义：数据集进行离线发布所使用的介质名称

英文名称：mediaName

数据类型：字符串

取值规则：介质名称代码<<代码表>>（见本标准 9.2.2）

短 名：medName

注 解：可选项；最大出现次数为 1

8.2.5 数据质量信息子集

定 义：提供数据集质量的总体评价信息

英文名称：DataQualityInfo

数据类型：复合型

短 名：DataQuaInfo

子 元 素：数据质量信息子集 =

数据志说明

注 解：可选项；最大出现次数为 N

8.2.5.1 数据志说明

定 义：数据集生产者对数据源和处理步骤的一般说明

英文名称：statement

数据类型：字符串

取值规则：自由文本

短 名：statement

注 解：必选项；最大出现次数为 1

8.2.6 限制信息子集

定 义：对数据集和元数据进行访问和使用的限制信息

英文名称：Data Constraints

数据类型：复合型

短 名：DataConst

子 元 素：限制信息子集 =
 $0\{\text{数据集限制信息}\}n+$
 $0\{\text{元数据限制信息}\}n+$

注 解：可选项；最大出现次数为 N

8.2.6.1 数据集限制信息

定 义：提供数据集和元数据进行访问和使用的限制信息

英文名称：dataset Constraints

数据类型：复合型

取值规则：限制信息<<数据类型>>（见本标准 9.1.2）

短 名：dataConst

注 解：可选项，最大出现次数 N

8.2.6.2 元数据限制信息

定 义：对元数据进行访问和使用的限制信息

英文名称：metadata Constraints

数据类型：复合型

取值规则：限制信息<<数据类型>>（见本标准 9.1.2）

短 名：mdConst

注 解：可选项，最大出现次数 N

8.2.7 维护信息子集

定 义：对数据集或元数据进行更新与维护的有关信息

英文名称：Maintenance Info

数据类型：复合型

短 名：MaintInfo

子 元 素：维护信息=
 $0\{\text{数据集维护信息}\}1+$
 $0\{\text{元数据维护信息}\}1$

注 解：可选项；最大出现次数为 1

8.2.7.1 数据集维护信息

定 义：对数据集进行更新与维护的有关信息

英文名称：dataset Maintenance Info

数据类型：复合型

取值规则：数据维护信息<<数据类型>>（见本标准 9.1.3）

短 名: dataMaint
注 解: 可选项, 最大出现次数为 1

8.2.7.2 元数据维护信息

定 义: 元数据更新与维护的相关信息
英文名称: metadataMaintenance
数据类型: 复合型
取值规则: 数据维护信息<< 数据类型>> (见本标准 9.1.3)
短 名: mdMaint
注 解: 可选项, 最大出现次数为 1

9 引用信息与代码表

9.1 引用信息

9.1.1 负责方<<数据类型>>

定 义: 负责单位或联系人的相关信息
英文名称: Responsible Party
数据类型: 复合型
短 名: RespParty
注 解: 可选项; 最大出现次数为 N
子 元 素: 负责方=

负责单位名称+
0{联系人姓名}n+
0{联系电话}n+
0{负责方地址} 1+
0{邮政编码} 1+
0{电子邮件地址} n

9.1.1.1 负责单位名称

定 义: 负责单位名称
英文名称: responsible Organization Name
数据类型: 字符串
取值规则: 自由文本
短 名: rpOrgName
注 解: 必选项; 最大出现次数为 1

9.1.1.2 联系人姓名

定 义: 联系人姓名
英文名称: responsible Individual Name
数据类型: 字符串
取值规则: 自由文本
短 名: rpIndName
注 解: 可选项; 最大出现次数为 N

9.1.1.3 联系电话

定 义: 可以与负责人或负责单位联系的电话号码
英文名称: contact Phone
数据类型: 字符串
取值规则: 自由文本
短 名: cntPhone

注 解：可选项；最大出现次数为 N

9.1.1.4 通讯地址

定 义：能够进行邮政联系的详细地址

英文名称：postal Address

数据类型：字符串

取值规则：自由文本

短 名：postAddr

注 解：可选项；最大出现次数为 1

9.1.1.5 邮政编码

定 义：邮政编码

英文名称：postal Code

数据类型：字符串

取值规则：按 GB/T 2260 的规定

短 名：postCode

注 解：可选项；最大出现次数为 1

9.1.1.6 电子邮件地址

定 义：联系人或负责单位的电子邮件地址

英文名称：electronic Mail Address

数据类型：字符串

取值规则：自由文本

短 名：eMailAdd

注 解：可选项；最大出现次数为 N

9.1.2 限制信息<<数据类型>>

定 义：对数据集和元数据的访问、使用以及安全施加的限制和约束

英文名称：Constraints

数据类型：字符串

短 名：Consts

子 元 素：限制信息 =

1 {访问限制|使用限制}n+

安全限制分级

注 解：条件必选项；最大出现次数为 N；当不选用“使用限制”时为必选

9.1.2.1 访问限制

定 义：为保护隐私权或知识产权，对访问元数据施加的限制和约束

英文名称：access Constraints

数据类型：字符串

取值规则：限制代码<<代码表>>（见本标准 9.2.3）

短 名：accessConsts

注 解：条件必选项；最大出现次数为 N；当不选用“使用限制”时为必选

9.1.2.2 使用限制

定 义：为保护隐私权或知识产权，对使用元数据施加的限制和约束

英文名称：use Constraints

数据类型：字符串

取值规则：限制代码<<代码表>>（见本标准 9.2.3）

短 名：useConsts

注 解：条件必选项；最大出现次数为 N；当不选用“访问限制”时为必选

9.1.2.3 安全限制分级

定 义：对元数据处理限制的名称
 英文名称：security Classification
 数据类型：字符串
 取值规则：安全限制分级代码<<代码表>>（见本标准 9.2.4）
 短 名：secuClass
 注 解：必选项；最大出现次数为 1

9.1.3 数据维护信息<<数据类型>>

定 义：对数据进行更新与维护的有关信息
 英文名称：Data Maintenance
 数据类型：复合型
 短 名：DataMaint
 子 元 素：数据维护信息 =
 维护和更新频率 +
 0{更新范围说明 }n +
 0{维护注释}n +
 0{维护方联系信息}n

9.1.3.1 维护和更新频率

定 义：在数据集初次完成后，对其进行修改和补充的频率
 英文名称：maintenance And Update Frequency
 数据类型：字符串
 取值规则：维护频率代码<<代码表>>（见本标准 9.2.5）
 短 名：mainFreq
 注 解：必选项；最大出现次数为 1

9.1.3.2 更新范围说明

定 义：数据集内容更新范围的说明
 英文名称：update Scope Description
 数据类型：字符串
 取值规则：范围代码<<代码表>>（见本标准 9.2.6）
 短 名：upScpDesc
 注 解：可选项；最大出现次数为 N

9.1.3.3 维护注释

定 义：有关对数据集维护更新的特殊需求信息
 英文名称：maintenance Note
 数据类型：字符串
 取值规则：自由文本
 短 名：maintNote
 注 解：可选项；最大出现次数为 N

9.1.3.4 维护方联系信息

定 义：联系负责维护数据集的人和单位的标识和方法
 英文名称：contact
 数据类型：复合型

取值规则：负责方<<数据类型>>（见本标准 9.1.1）

短 名：maintCont

9.2 代码表

9.2.1 语种代码表<<代码表>>

	名称（中文）	名称（英文）	域代码	定义
	语种代码	LanguageCode	LangCd	
1	zh-HK	zh-HK	0x0C04	中文 - 中国香港特别行政区
2	zh-MO	zh-MO	0x1404	中文 - 中国澳门特别行政区
3	zh-CN	zh-CN	0x0804	中文 - 中国
4	zh-CHS	zh-CHS	0x0004	中文（简体）
5	zh-SG	zh-SG	0x1004	中文 - 新加坡
6	zh-TW	zh-TW	0x0404	中文 - 中国台湾
7	zh-CHT	zh-CHT	0x7C04	中文（繁体）

9.2.2 介质名称代码<<代码表>>

	名称（中文）	名称（英文）	域代码	定义
	介质名称代码	MediumNameCode	MedNameCd	介质名称
1	只读光盘	cdrom	001	只读光盘
2	数字视频光盘	dvd	002	数字视频光盘
3	数字视频只读光盘	dvdRom	003	数字视频只读光盘
4	3'软盘	3halfInchFloppy	004	3'软盘
5	5'软盘	5quarterInchFloppy	005	5'软盘
6	7 磁道磁带	7trackTape	006	7 赤道磁带
7	9 磁道磁带	9rackTape	007	9 磁道磁带
8	3480 盒式磁带	3480 Cartridge	008	3480 盒式磁带
9	3490 盒式磁带	3490 Cartridge	009	3490 盒式磁带
10	3580 盒式磁带	3580 Cartridge	010	3580 盒式磁带
11	4mm 盒式磁带	4mm Cartridge	011	4mm 盒式磁带
12	8mm 盒式磁带	8mm Cartridge	012	8mm 盒式磁带
13	1/4' 盒式磁带	1quarterInchCartridgeTape	013	1/4' 盒式磁带
14	数字线形磁带	digitalLinearTape	014	半英寸数据流盒式磁带
15	在线	onLine	015	直接连接计算机
16	卫星	satellite	016	通过卫星通信系统连接
17	电话连接	telephoneLink	017	通过电话网通信
18	硬拷贝	hardcopy	018	提供说明信息的手册或简介

9.2.3 限制代码<<代码表>>

	名称（中文）	名称（英文）	域代码	定义
	限制代码	RestrictionCode	RestrictCd	对访问或使用数据施加的限制
1	版权	copyright	001	法律批准的发行者在确定的时间内，对出版、创作或销售数据资料的专有权利，或使用商业印刷品或商标的权利
2	专利权	patent	002	政府已经批准的制造、出售、使用或特许发明或发现的专门权利
3	专利审查权	patentPending	003	等待专利权的生产或销售信息
4	商标	trademark	004	正式注册标识产品的、法律上只限于所有者或厂商使用的名称、符号或其他图案
5	许可证	license	005	正式许可作某事
6	知识产权	intellectualPropertyRights	006	从创造活动生产的无形资产的分发或分发控制获得经济的权利
7	受限制	restricted	007	控制一般的流通或公开
8	其他限制	otherRestrictions	008	未列出的限制

9.2.4 安全限制分级代码<<代码表>>

	名称（中文）	名称（英文）	域代码	定义
	安全限制分级代码	ClassificationCode	ClasscationCd	对于数据集操作进行限制的名称
1	完全公开	open	001	国内外公开
2	专业公开		002	专业人员内部公开
3	内部公开	restricted	003	项目组内部公开
4	秘密	confidential	004	受委托者可以使用该信息

9.2.5 维护频率代码<<代码表>>

	名称（中文）	名称（英文）	域代码	定义
	维护频率代码	Maintenance FrequencyCode	MaintFreqCd	在数据集第一次生成后，对其进行修改和删除的频率
1	连续	continual	001	数据重复地频繁地进行更新
2	按日	daily	002	数据每天更新一次
3	按周	weekly	003	数据每周更新一次
4	按两周	fornightly	004	数据每两周更新一次
5	按月	monthly	005	数据每月更新一次
6	按季	quarterly	006	数据每季更新一次
7	按半年	biannually	007	数据每年更新两次
8	按年	annually	008	数据每年更新一次
9	按需要	asNeeded	009	数据需要更新
10	不固定	irregular	010	数据不定期更新
11	无计划	notPlanned	011	尚无更新计划
12	未知	unknown	012	数据维护频率未知

9.2.6 范围代码<<代码表>>

	名称（中文）	名称（英文）	域代码	定义
	范围代码	ScopeCode	ScopeCd	数据集内容更新范围的说明
1	全部	whole	001	数据集全部内容被更新
2	大部分	mostly	002	数据集大部分内容被更新
3	部分	part	003	数据集部分内容被更新
4	少量	little	004	数据集少量内容被更新

附录 A

(资料性)

卫生健康信息数据集核心元数据著录示例

以《2022年中国卫生健康统计年鉴》中的卫生设施数据集为例，进行卫生健康信息数据集的核心元数据著录。

《2022年中国卫生健康统计年鉴》中，包含了医疗卫生机构、卫生人员、卫生设施、卫生经费、医疗服务、基础医疗卫生服务、中医药服务、妇幼保健、人民健康水平、疾病控制与公共卫生、居民病伤死亡原因等16个部分，每个部分围绕主题内容进行收据的收集、形成数据集，分别构成16个数据集。以其中的卫生设施数据及内容为例，进行核心元数据的著录见表A.1。

表 A.1 2022 年中国卫生健康统计年鉴卫生设施数据集核心元数据著录示例

元数据项		元数据值
数据集名称		2022年中国卫生健康统计年鉴卫生设施数据集
数据集标识符		国家卫生健康委员会-2022年中国卫生健康统计年鉴-卫生设施-1.0
数据集摘要		主要介绍全国及31个省、自治区、直辖市医疗卫生机构床位、医用设备、房屋面积和信息化基础资源情况。主要包括各级各类医疗卫生机构床位数；医院、社区卫生服务中心、乡镇卫生院主要医用设备台数，各类医疗卫生机构房屋建筑面积、公立医院信息化基础资源情况等。
数据集提交或发布方	负责单位名称	国家卫生健康委统计信息中心
	联系人姓名	
	联系电话	
	通讯地址	北京市西城区西直门外南路1号
	邮政编码	100044
关键词说明	电子邮件地址	
	关键词1	卫生设施
	关键词2	卫生统计
词典名称		
特征数据元		床位数，每千人口医院，卫生院床位数，设备台数，购建房屋建筑面积
数据集发布日期		2022-07-01
数据集发布格式	发布格式名称	
	版本	
相关环境说明		《2022年中国卫生健康统计年鉴》内容的一个部分，以卫生设施为主题的数据集合。
元数据创建日期		2022-07-15
元数据标识符		MD-国家卫生健康委员会-2022年中国卫生健康统计年鉴-卫生设施-1.0
元数据负责方	负责单位名称	国家卫生健康委统计信息中心
	联系人姓名	
	联系电话	
	通讯地址	
	邮政编码	
电子邮件地址		
元数据标准名称		卫生健康信息数据集元数据标准
元数据标准版本		
元数据更新日期		

附录 B

(规范性)

卫生健康信息数据集元数据扩展原则与方法

在卫生健康领域针对数据集制定专用元数据标准时，如果在本标准的参考元数据中没有可以直接使用的元数据子集/实体/元素/代码表等时，可以依据本节中的原则和方法来扩展新的元数据子集/实体/元素/代码表等。

B.1 元数据扩展原则

新扩展的元数据需要遵循如下两条基本原则：

- a) 新建的元数据不应与共享参考元数据中的现有的元数据子集、实体、元素、代码表的名称、定义相冲突。
- b) 如果共享元数据标准中规定了不允许的情况，则不得扩展这些不允许的任何内容。

B.2 元数据扩展的步骤

扩展元数据时，步骤见图B.1。

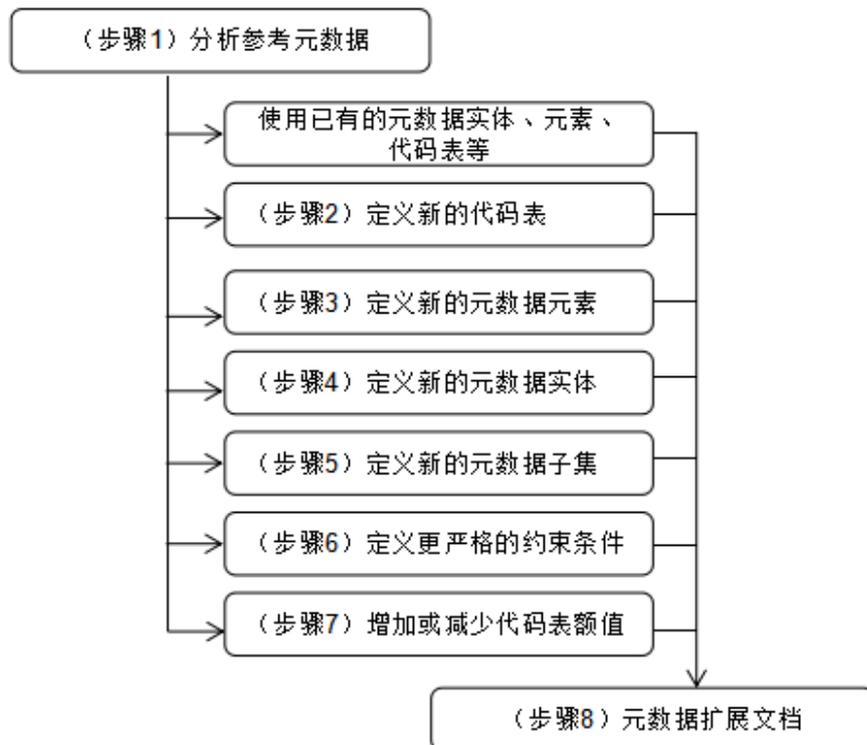


图 B.1 元数据扩展的步骤

B.2.1 步骤 1 分析共享参考元数据

扩展的第 1 步应保证对共享参考元数据进行全面的分析，这种分析不仅要针对元数据实体/元素的名称，还应分析它们的定义、数据类型、约束条件、值域和最大出现次数等属性，在不能满足需要的情况下进行扩展。方法如下：

- a) 如果共享参考元数据中存在能够满足要求的元数据子集、元素、实体、代码表，则直接采用

- 即可，无需新扩展元数据。
- b) 在共享参考元数据中的元数据代码表无法满足领域要求的情况下，需要通过扩展新的元数据代码表以满足需要，则进入步骤 2。
 - c) 在共享参考元数据中的元数据元素无法满足领域要求的情况下，需要通过扩展新的元数据元素以满足需要，则进入步骤 3。
 - d) 在共享参考元数据中的元数据实体无法满足领域要求的情况下，需要通过扩展新的元数据实体以满足需要，则进入步骤 4。
 - e) 在共享参考元数据中的元数据子集无法满足领域要求的情况下，需要通过扩展新的元数据实体以满足需要，则进入步骤 5。
 - f) 通过更改共享参考元数据中的元数据的约束条件就可以满足要求的情况下，则进入步骤 6。
 - g) 在共享参考元数据中代码表的值需要扩展的情况下，则进入步骤 7。

B.2.2 步骤 2 定义新的代码表

在需要一个新的代码表以满足某个元数据元素值域需要时，需要做以下工作：

- a) 建立新的元数据代码表，并添加代码表中的值；
- b) 进入步骤 8，建立元数据扩展文档。
- c) 使用新元数据代码表以满足需求。

B.2.3 步骤 3 定义新的元数据元素

在需要一个新的元数据元素以满足需要时，需要做以下工作：

- a) 给出新元数据元素的中文名称、英文名称、定义、数据类型、值域、短名、约束/条件、最大出现次数等属性信息。
- b) 如果它需要新的代码表，则进入步骤 2。
- c) 进入步骤 8，建立元数据扩展文档。
- d) 使用新元数据元素以满足需求。

B.2.4 步骤 4 定义新的元数据实体

在需要一个新的元数据实体以满足需要时，需要做以下工作：

- a) 给出新元数据实体的中文名称、定义、英文名称、数据类型、短名、子元素、注解等属性信息。
- b) 确定构成元数据实体的元数据元素。
- c) 如果构成该元数据实体的元数据元素需要新建，则进入步骤 3。
- d) 进入步骤 8，建立元数据扩展文档。
- e) 使用新元数据实体以满足需求。

B.2.5 步骤 5 定义新的元数据子集

在需要一个新的元数据子集以满足需要时，需要做以下工作：

- a) 建立新的元数据子集，确定该元数据子集的定义。
- b) 确定新建的元数据子集中包含的元数据元素、实体等。
- c) 新元数据子集中包含的元数据元素需要新建，则进入步骤 3。
- d) 新元数据子集中包含的元数据实体需要新建，则进入步骤 4。
- e) 进入步骤 8，建立元数据扩展文档。
- f) 使用新元数据子集中的内容以满足需求。

B.2.6 步骤 6 定义更严格的元数据约束条件

需要选用一个共享参考元数据中的已有的元数据实体、元素，但需要其具备更严格的约束条件，则可以用“必选”代替“条件必选”或“可选”，可以用“条件必选”代替“可选”。方法是：

- a) 定义该元数据实体、元素新的约束条件。如果新的条件约束是“条件必选”，则应给出必须使用该元数据实体、元素时的条件
- b) 进入步骤 8，建立元数据扩展文档。
- c) 使用新的元数据。

B.2.7 步骤 7 增加或减少代码表的值

需要选用一个共享参考元数据中的代码表，但需要通过减少或增加代码表中的项来对原有的代码表进行特化，方法是：

- a) 修改该代码表，较少或增加相应的项。
- b) 进入步骤 8，建立元数据扩展文档。
- c) 使用新的代码表。

B.2.8 步骤 8 元数据扩展文档

一旦定义了新元数据子集、实体、元素，需要明确地记录对共享元数据的改变，这种改变必须按相应格式在新标准文档中记录，具体方法如下：

- a) 如果建立的是新的代码表，则要增加新的的代码表信息。
 - b) 如果建立的是新的元数据元素、实体或子集，则要更新对应的摘要描述信息。
 - c) 如果定义更严格的元数据约束条件，要更新相应的摘要描述信息。
 - d) 增加或减少代码表的值，则要更新相应的代码表。
-