

科学数据共享工程技术标准

SDS/T 2113—2004

元数据的 XML/XSD 置标规则

XML/XSD marking rules for metadata

(征求意见稿)

(本稿完成日期: 2005 年 4 月)

××××-××-××发布

××××-××-××实施

中华人民共和国科学技术部 发布

目 次

目 次.....	I
前 言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和缩略语.....	1
4 总体思路.....	2
5 元数据规范的 UML 建模规则	3
6 元数据的 UML 模型向 XML 模式映射的规则	4

前 言

元数据是对信息资源的规范化描述，它是按照一定标准，从信息资源中抽取相应的特征，组成的一个特征元素集合。这种规范化描述可以准确和完备地说明信息资源的各项特征。

元数据内容标准一般包括了完整描述一个具体对象时所需要的数据项集合、各数据项语义定义、著录规则和计算机应用时的语法规则等。元数据内容标准要从功能、数据结构、格式、语义、语法等角度来进行标准的制定。

当要求以自动化或无人（或最少的人工）干预的形式在同构或异构系统间进行元数据内容的交换和自动处理时，就需要依托 XML 语法规则，按照一种一致的方式，将各种面向科学数据共享的元数据用 XML 模式来定义，以便于各试点以统一的方式对科学数据进行存储、交换和共享。

按照向上积极采用适宜的国际标准或国家标准，向下突出科学数据共享活动的特点和需求的原则，本标准结合科学数据共享平台的实际情况，对标准中所述方法和规则进行了验证。

本标准第一次制定。

本标准由中华人民共和国科学技术部基础研究司提出。

本标准由中华人民共和国科学技术部基础研究司归口。

本标准主要起草单位：中国标准化研究院。

本标准主要起草人：刘颖等（待补充）

元数据的 XML/XSD 置标规则

1 范围

本标准规定了科学数据共享平台建设中，将元数据用 XML 模式来定义的方法和规则。

本标准适用于所有需要将各种基于元数据标准的科学数据用 XML 模式来定义的场所。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本标准。然而，鼓励根据本标准达成协议的各方，研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

SDS/T××××—2004 科学数据共享 元数据标准化的基本原则与方法

SDS/T××××—2004 科学数据共享 核心元数据

SDS/T××××—2004 科学数据共享 元数据内容

SDS/T××××—2004 科学数据共享 元数据检索和提取协议

农村科技数据共享平台 元数据规范（V1.0 版）

GB/T 18793-2002 信息技术 可扩展置标语言（XML）1.0

W3C XML Schema Part 0: Primer

W3C XML Schema Part1: Structures

W3C XML Schema Part2: Datatypes

W3C Namespaces in XML

3 术语和缩略语

本标准采用下列术语和定义。

3.1

数据类型 data type

值域说明，允许对该值域内的值进行操作。如：字符串(string)、十进制(decimal)、整形(integer)、布尔型(boolean)、日期型(date)和二进制(binary)。

3.2

元数据 metadata

关于数据的数据。

3.3

元数据元素 metadata element

元数据的基本单元。

3.4

元数据实体 metadata entity

一组说明数据相同特性的元数据元素。

3.5

元数据根实体 metadata root entity

3.6

元数据子集 metadata section

元数据的子集合，由相关的元数据实体和元素组成。

3.7

元数据类型实体 metadata type entity

3.8

XML 命名空间 XML namespace

为了解决命名冲突，为元素和属性命名引入的逻辑空间，是在 XML 文档中通过 URI 引用声明的，并采用限定性前缀将元素和属性与命名空间联系起来。

3.9

XML 模式 XML Schema

指出哪些元素允许出现在 XML 文档中以及它们以何种方式组合的元素名称的形式规范。它定义文档的结构，如哪些元素是其他元素的子元素，子元素出现的顺序以及子元素的数目。它还可以定义元素为空还是包括文本，同时还能定义属性的缺省值。本标准中的 XML 模式特指 W3C 制定的 Schema 标准（REC-xmlschema-0-20010502 Part 0-2）。

4 总体思路

为了让用户能以一种一致的方式将元数据内容用XML模式来定义，比较好的办法是通过一种模型化方法，首先将元数据抽象为一种独立于语法的、中性的模型（推荐使用UML模型），这种模型可以与XML语法绑定，然后按照转换规则形成XML模式。因此，总体思路按照以下两大步骤进行：

步骤1：首先按照科学数据共享《元数据内容》标准中提供的方法和原则，将元数据内容用UML的模型进行抽象，形成元数据的UML模型，这种模型是与语法无关的、中性的。

步骤2：在元数据的UML模型基础之上，依据W3C XML模式的语法规则，将元数据的UML模型中元数据实体、元数据类型实体、元数据元素以及元数据中的代码表映射成XML模式。

5 元数据规范的 UML 建模规则

有关元数据内容以及相关的 UML 内容参见科学数据共享《元数据内容》。

本章根据第 4 章中所述步骤 1，提供了较为详细的、具体的将元数据抽象为 UML 模型的方法和规则。

5.1 元数据组成结构和表达格式

根据第 3 章术语中的元数据定义，可以将元数据理解为：元数据是描述一个具体的资源对象，并能对该对象进行定位、管理，且有助于它的发现与获取的数据。

元数据在具体应用时，体现为描述某类资源的具体对象的元数据子集。元数据子集由许多完成不同功能的元数据实体和元数据元素组成；元数据实体是由元数据元素组成的，元数据实体可以与一个或多个元数据类型实体相关联；元数据实体和元数据元素可以用数据字典的方式描述，并通过以下属性定义元数据实体和元数据元素，即：名称/角色的中文名称、名称/角色的英文名称、短名、定义、约束/条件、最大出现次数、数据类型、域。

图 1 所示为元数据子集、元数据实体、元数据类型实体和元数据元素间的关系。

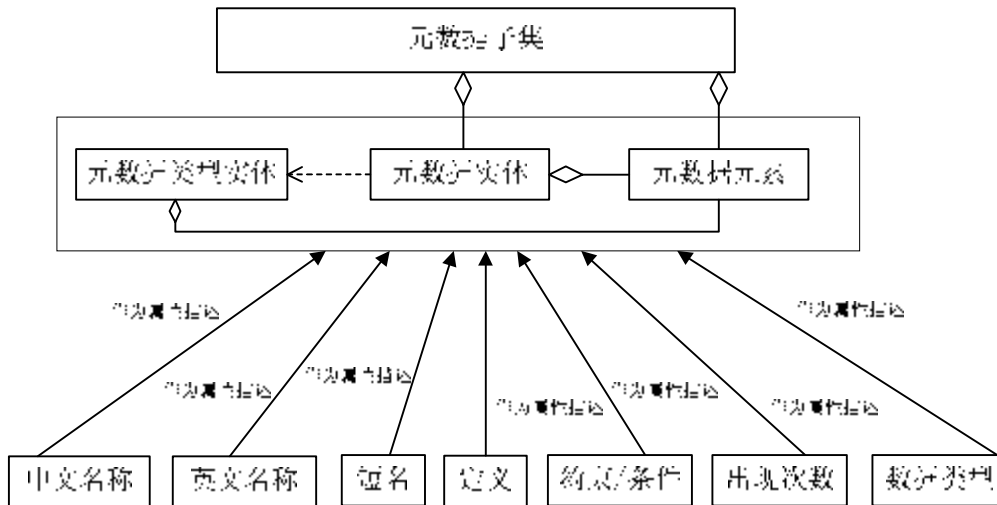


图 1 元数据子集、元数据实体、元数据类型实体、元数据元素以及描述属性之间的关系

5.2 UML 类图及其组成结构

用 UML 的静态结构图建立元数据内容模型，用 UML 类的概念表示元数据实体和元数据类型实体，类与类间的关系表示元数据实体间的关系。

UML 类的组成结构和表达格式如图 2 所示。

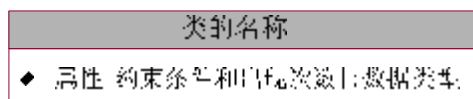


图 2 类的组成结构和表达格式

5.3 元数据转换为 UML 模型的规则

将 5.1 中描述的元数据表达格式和 5.2 中描述的 UML 类的表达格式相对应，建立如图 3 所示的

元数据与 UML 模型的对应关系。

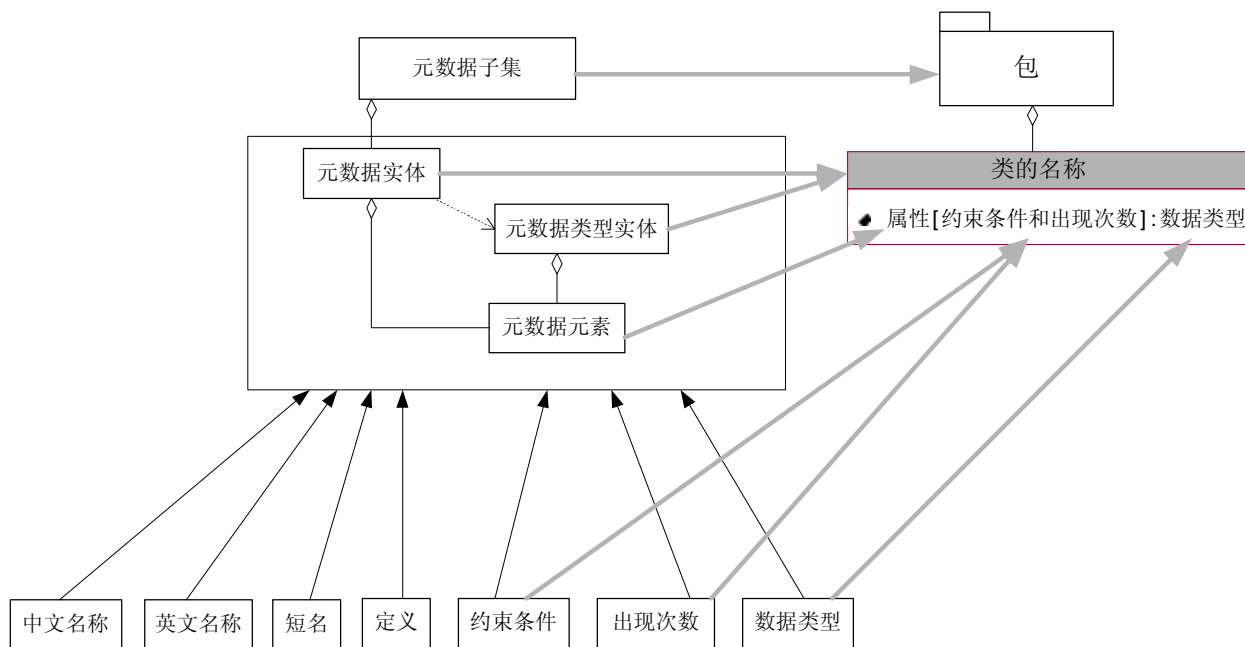


图 3 元数据与 UML 模型的对应关系

因此，确定的元数据转换为 UML 模型的规则如下：

规则 1: 应将描述资源具体对象的元数据子集转化为 UML 模型中的包，元数据子集的名称对应于 UML 模型中包的名称。

规则 2: 应将元数据实体和元数据类型实体转化为 UML 模型中的类，这些类可以是特化的（子类）或泛化的（超类）。

注：元数据实体可以与一个或多个元数据类型实体相关联。

规则 3: 应将元数据实体所包含的元数据元素转化为 UML 模型中该元数据实体所对应类中的属性。

规则 4: 应将描述元数据元素的约束条件和出现次数转化为类中属性的约束条件和出现次数。

元数据元素的约束条件和出现次数与类中属性的约束条件和出现次数的对应关系如表 1 所示：

表 1：元数据元素的约束条件和出现次数与类中属性的约束条件和出现次数的对应关系

元数据元素的约束条件和出现次数	类中属性的约束条件和出现次数
可选，最大出现次数为 1 次	0..1
可选，最大出现次数为 N 次	0..N
必选，最大出现次数为 1 次	1..1
必选，最大出现次数为 N 次	1..N

规则 5: 应将描述元数据元素的数据类型转化为类中属性的数据类型。

实例：参见《科学数据共享元数据》标准中图 A-1、A-2。

6 元数据的 UML 模型向 XML 模式映射的规则

本章规定了一套较为详细的规则，按照这套规则可以将元数据的 UML 模型与 XML 语法定义，然后以一种一致的方式将模型转换为 XML 模式。

6.1 总则

规则 1: 应详细说明 XML 模式的前导说明 (prolog) 部分;

规则 2: 应将元数据根实体定义为一个根元素;

规则 3: 应将元数据实体定义为 XML 模式的元素 (element);

规则 4: 应将元数据类型实体定义为 XML 模式的复杂类型 (complexType);

规则 5: 应将元数据元素定义为 XML 模式的元素 (element);

规则 6: 应将元数据标准中的代码表定义为 XML 模式的简单类型 (simpleType);

规则 7: 元数据子集不应转换为 XML 模式。

规则 8: 元数据的 XML 模式应符合 W3C XML 模式的建议书: 第 1 部分: 结构和第 2 部分: 数据类型。

元数据与 XML 模式的映射关系如图 5 所示。

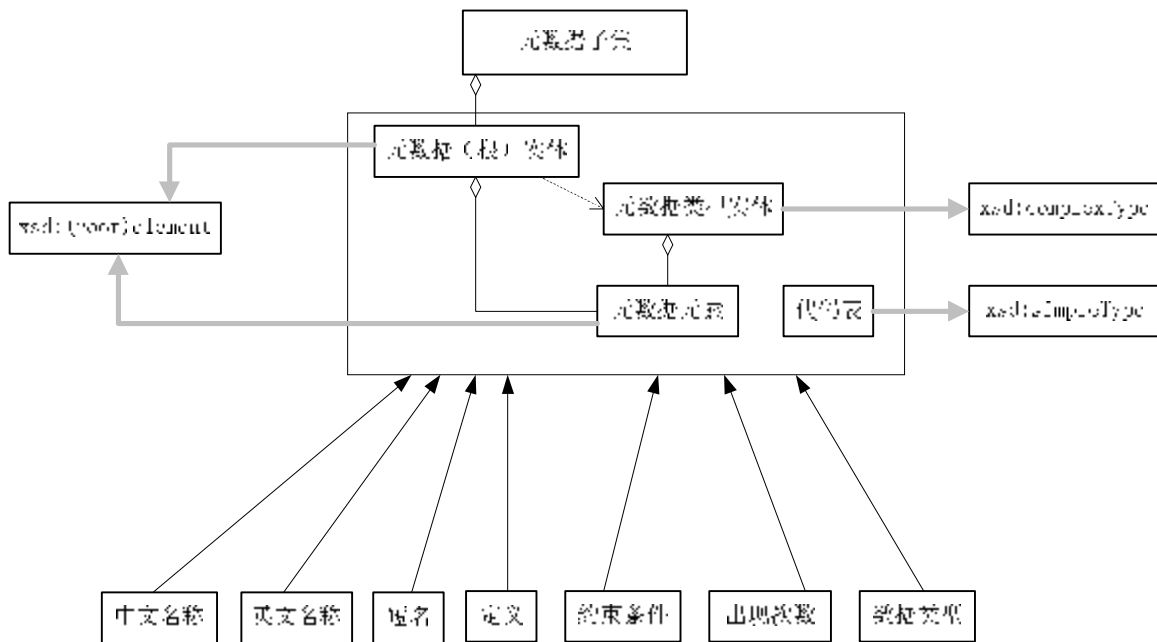


图 5 元数据与 XML 模式的映射关系

6.2 XML 模式的前导说明部分 (prolog)

规则9: 字符集的说明

- 1、如果XML模式中元素、属性、数据类型等内容都用英文表达，则encoding属性值为"UTF-8";
- 2、如果XML模式中元素、属性、数据类型等内容中包含中文，则encoding属性值为"GB-2312"。

规则10: 命名空间 (namespace) 的说明

- 1、XML 模式的前导说明部分必须包含命名空间的定义，缺省命名空间的值为“http://sciencedata.cn”；

实例：

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="http://sciencedata.cn"
xmlns:scimd="http://sciencedata.cn"
```

- 2、科学数据共享元数据内容的命名空间的值为“http://sciencedata.cn”；
- 3、领域科学数据共享元数据内容的命名空间的值为“http://XXX.sciencedata.cn”；

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="http://sciencedata.cn"
xmlns:scimd=http://sciencedata.cn
xmlns:   =http://XXX.sciencedata.cn
```

- 4、领域专用科学数据共享元数据内容的命名空间的值为“http://YYY.XXX.sciencedata.cn”；

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="http://sciencedata.cn"
xmlns:scimd=http://sciencedata.cn
xmlns:   =http://XXX.sciencedata.cn
xmlns:   =http://YYY.XXX.sciencedata.cn
```

注：XXX内容由国家科学数据中心或国家科学数据网确定，YYY内容由国家科学数据中心分中心或国家科学数据网的分节点确定。（加术语）

- 5、应在前导说明部分以注释的形式说明以下内容：

- 元数据标准名称；
- 元数据标准版本；
- XML模式编写单位或编写人；
- XML模式完成时间；

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="http://sciencedata.cn"
xmlns:scimd="http://sciencedata.cn"
```

6.3 元数据根实体定义为一个根元素

规则11：元数据的根实体定义为XML模式的根元素，根元素的名称是“metadata”，根元素的类型定义为“xs:complexType”。

例如：声明“科学数据共享元数据”作为XML模式的根元素为：

```
<xs:element name="科学数据共享元数据" type="科学数据共享元数据类型">
  <xs:annotation>
```

```

    <xs:documentation>根元素</xs:documentation>
  </xs:annotation>
</xs:element>
<xs:complexType name="科学数据共享元数据类型">
  <xs:sequence>
    .....
  </xs:sequence>
</xs:complexType>

```

6.4 元数据实体定义为 XML 模式的元素 (element)

规则 12: 元数据实体名称定义为 XML 模式中的 element name, element name 的值必须用元数据实体的短名。

规则 13: 元数据实体中文名称定义为该元数据实体对应元素的属性, 即:

```
<xs:attribute name="元数据实体中文名称" type="xs:string" use="required"/>
```

规则 14: 元数据实体英文名称定义为该元数据实体对应元素的属性, 即:

```
<xs:attribute name="元数据实体英文名称" type="xs:string" use="required"/>
```

规则 15: 元数据实体的类型定义为 element name 所包含的 complexType。如: 根据《科学数据共享元数据内容》标准 0.9 版中的 B.1.2.1, 元数据实体“标识信息”的 XML 模式代码如下:

```

<xs:element name="Ident">
  <xs:annotation>
    <xs:documentation>元数据描述的资源的基本信息</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element name="idCitation" type="extent">
        <xs:annotation>
          <xs:documentation>数据集引用信息</xs:documentation>
        </xs:annotation>
      </xs:element>
      <xs:element name="idAbs" type="string">
        <xs:annotation>
          <xs:documentation>数据集内容的简单说明</xs:documentation>
        </xs:annotation>
      </xs:element>
      .....
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:complexType name="extent">
  <xs:annotation>
    <xs:documentation>平面、垂向何时间覆盖范围信息</xs:documentation>
  </xs:annotation>
  <xs:sequence>

```

```

    <xs:element name="exDesc" type="string">
      <xs:annotation>
        <xs:documentation>有关对象的空间和时间覆盖范围</xs:documentation>
      </xs:annotation>
    </xs:element>
    .....
  </xs:sequence>
</xs:complexType>

```

规则 16: 元数据实体的定义、注解用注释说明, 即:

```

<xs:annotation>
  <xs:documentation>元数据实体的定义、注解等</xs:documentation>
</xs:annotation>

```

规则 17: 元数据实体的出现次数用 `maxOccurs` 和 `minOccurs` 表达。

其中 `maxOccurs` 表示该元数据实体可以具有的最大实例数目。

当元数据实体是可选项或条件必选项时, 用 `minOccurs` 表示该元数据实体的可选属性, 且 `minOccurs=0`。

当元数据实体是必选项时, `minOccurs=1`。

规则 18: 如果组成元数据实体的元数据元素之间、元数据实体之间、元数据元素和元数据实体之间有条件必选关系, 使用“`xs:choice`”表示, 如果没有条件必选关系, 只是序列关系, 则用 `sequence` 表示。

如: 科学数据共享《元数据标准》0.9 版中, 附录 B.2.2 元数据实体“负责方单位信息”包含的 3 个子元素“负责人姓名、负责人单位名称和负责人职务”之间是条件必选关系, 使用“`xs:choice`”表示为:

```

<xs:complexType>
  <xs:choice>
    <xs:element name="rpIndName" type="xs:string"/>
    <xs:element name="rpOrgName" type="xs:string"/>
    <xs:element name="rpPosName" type="xs:string"/>
  </xs:choice>
</xs:complexType>

```

6.5 元数据类型实体定义为 XML 模式的复杂类型 (complexType)

规则 19: 元数据类型实体定义为 `complexType`, `complexType name` 的值必须用元数据类型实体的短名。如: 在科学数据共享《元数据标准》0.9 版中, 附录 B2.2.2 中“引用和负责单位信息”是元数据类型实体, 按照本规则, 相应的 XML 模式代码为:

```

<xs:complexType name="Citation">
  <xs:sequence>
    <xs:element name="resTitle">
    <xs:element name="resRefDate">
    .....
  </xs:sequence>
</xs:complexType>

```

规则 20: 元数据类型实体中文名称定义为该元数据类型实体对应的 complexType name 的属性, 即:

```
<xs:attribute name="元数据类型实体中文名称" type="xs:string" use="required"/>
```

规则 21: 元数据类型实体英文名称定义为该元数据类型实体对应的 complexType name 的属性, 即:

```
<xs:attribute name="元数据类型实体英文名称" type="xs:string" use="required"/>
```

规则 22: 元数据类型实体的定义、注解用注释说明。

```
<xs:annotation>
```

```
<xs:documentation>元数据类型实体的定义、注解等</xs:documentation>
```

```
</xs:annotation>
```

6.5 元数据元素定义为 XML 模式的元素 (element)

规则 23: 元数据元素名称定义为 element name, element name 的值必须用元数据元素的短名。

规则 24: 元数据元素中文名称定义为该元数据元素对应元素的属性, 即:

```
<xs:attribute name="元数据元素中文名称" type="xs:string" use="required"/>
```

规则 25: 元数据元素英文名称定义为该元数据元素对应元素的属性, 即:

```
<xs:attribute name="元数据元素英文名称" type="xs:string" use="required"/>
```

规则 26: 元数据元素的类型用 type 属性说明, type 属性的值可以是元数据类型实体名称、元数据代码表名称或系统默认的数据类型名称。

规则 27: 元数据元素的定义、注解用注释说明。

```
<xs:annotation>
```

```
<xs:documentation>元数据类型实体的定义、注解等</xs:documentation>
```

```
</xs:annotation>
```

规则 28: 元数据元素的出现次数用 maxOccurs 和 minOccurs 表达。

其中 maxOccurs 表示该元数据元素可以具有的最大实例数目。

当元数据元素是可选项或条件必选项时, 用 minOccurs 表示该元数据元素的可选属性, 且 minOccurs=0。

当元数据元素是必选项时, minOccurs=1。

规则 29: 元数据元素的值域, 没有特定约束无需说明。

如需说明, 则:

——对数值型类型的元数据元素, 用 “xs:maxLength” 和 “xs:minLength” 说明其最大最小范围;

——对文本型类型的元数据元素, 用 “xs:length” 说明其字符串长度。

6.6 元数据标准中的代码表定义为 XML 模式的简单类型 (simpleType)

规则 30: 元数据标准中的代码表定义为 simpleType, simpleType name 的取值必须用代码类型的域代码。

规则 31: 元数据代码表中文名称定义为该元数据代码表对应 “simpleType name” 的属性, 即:

```
<xs:attribute name="元数据代码表中文名称" type="xs:string" use="required"/>
```

规则 32: 元数据代码表英文名称定义为该元数据代码表对应“simpleType name”的属性，即：

```
<xs:attribute name="元数据代码表英文名称" type="xs:string" use="required"/>
```

规则 33: 元数据代码表的类型用 type 说明，type 的值为 string。

规则 34: 元数据代码表的定义用注释说明。

```
<xs:annotation>
```

```
  <xs:documentation>元数据代码表的定义</xs:documentation>
```

```
</xs:annotation>
```

规则 35: 元数据代码表的值定义为 simpleType 的枚举值。

规则 36: 元数据代码表中代码值的中文名称定义为该代码值对应“simpleType name”的属性，即：

```
<xs:attribute name="代码值中文名称" type="xs:string" use="required"/>
```

规则 37: 元数据代码表中代码值英文名称定义为该代码值对应“simpleType name”的属性，即：

```
<xs:attribute name="代码值英文名称" type="xs:string" use="required"/>
```

规则 38: 元数据代码表代码值的定义用注释（annotation）说明，即：

```
<xs:simpleType name="xx">
```

```
  <xs:restriction base="xs:string">
```

```
    <xs:enumeration value="ccc">
```

```
      <xs:annotation>
```

```
        <xs:documentation source="代码值ccc的定义"/>
```

```
      </xs:annotation>
```

```
    </xs:enumeration>
```

```
    <xs:enumeration value="bbb">
```

```
      <xs:annotation>
```

```
        <xs:documentation source="代码值bbb的说明"/>
```

```
      </xs:annotation>
```

```
    </xs:enumeration>
```

```
    <xs:enumeration value="aaa"/>
```

```
  </xs:restriction>
```

```
</xs:simpleType>
```