

# 中华人民共和国国家标准

GB/T 20533—2006

## 生态科学数据元数据

Metadata for ecological data

2006-09-18 发布

2007-03-01 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

## 前 言

本标准的附录 A、附录 B、附录 C、附录 D 和附录 E 为规范性附录，附录 F、附录 G 和附录 H 为资料性附录。

本标准由中国科学院“科学数据库及其应用系统”和“中国生态系统研究网络”支持编制。

本标准由中国科学院提出。

本标准由全国信息分类编码标准化技术委员会归口。

本标准起草单位：中国科学院计算机网络信息中心、中国科学院地理科学与资源研究所、中国科学院南京土壤研究所、中国科学院·水利部水土保持研究所、中国科学院寒区旱区环境与工程研究所、中国科学院·水利部成都山地灾害与环境研究所、中国科学院东北地理与农业生态研究所。

本标准主要起草人：黎建辉、肖云、于贵瑞、侯艳飞、胡良霖、施建平、何洪林、董济农、郭明航、虞路清、张耀南、陈斌如、赵军、王闰强、刘宁、高美荣。

本标准为第一次制定。

## 引 言

伴随着人类对人与自然协调发展重要性认识的加深,生态学受到越来越多的关注,并成为当前科学研究的热点学科。现代科学研究的一个特点是数据成为科学研究和科学发现的重要基础,在生态学领域亦是如此。

随着生态科学研究的发展和信息技术的进步,生态科学数据在迅速增长。然而目前与此并存的另一现象是,科研人员在实际的研究工作中往往感觉到所需数据的匮乏。很多时候,这种匮乏并不是因为所需的数据尚不存在,而在于这些数据难以发现、难以获取和缺乏可用性。这也是我国及其他很多国家大力开展和推进科学数据共享活动的缘起。在造成上述不协调现象的诸多因素中,元数据的缺乏和不规范是制约科研人员发现、管理和使用生态科学数据的主要原因之一。

元数据是关于数据的数据,用于说明数据的内容、品质、产生过程和背景、访问和获取方式以及其他有关特征。从数据的使用来看:

- a) 数据生产者可能在数据产生很长时间之后去使用或重用它们;
- b) 数据的生产者并非总是数据的使用者,数据生产者以外的人员使用数据的需求日益强烈,尤其是对长时间序列、大尺度和综合性生态研究来说,这样的需求更显而易见。

适当的元数据能帮助数据生产者以及数据生产者以外的用户更快地发现所需要的数据,更好地了解其内容和限制,评估其对于应用需求的适用性,并恰当地获取和使用它们。同时,由于数据生产者和用户都需要处理越来越多的数据,元数据能为他们提供关于这些数据的关键知识,帮助其有效地保存、管理和维护这些数据,提高效率,且使数据生产单位能够不受人员变动的影响,防止数据资产的流失,维护单位和国家的有关投资。

另外,在生态科学数据中,很多数据特别是观测数据的产生具有不可重复性,为保证它们不被误用和在将来可用,尤其需要详细地说明数据产生的方法和条件。元数据允许数据生产者对这些信息进行完全的记录,以便这些数据不因时间的流逝而丧失可用性。

本标准的目的是提供一个可扩展的、用于描述和归档管理生态科学数据的概念数据模型。该模型将生态科学数据的特征归纳为 10 类:标识信息、数据质量信息、方法信息、场地信息、项目信息、分发信息、元数据参考信息、实体信息、空间参照系信息和空间表示信息。本标准的实施将促进生态科学数据的高效保存、管理和维护,促进它们为科研人员、其他用户或者软件程序等的检索、评价、获取和使用,促进生态科学数据的共享和交换。

## 目 次

前言 .....	I
引言 .....	II
1 范围 .....	1
2 规范性应用文件 .....	1
3 术语和定义 .....	1
4 符号和缩略语 .....	4
4.1 缩略语 .....	4
4.2 UML 模型符号 .....	4
4.3 UML 模型关系 .....	5
4.4 UML 模型构造型 .....	5
5 一致性 .....	6
6 约定 .....	6
6.1 元数据基本属性 .....	6
6.2 UML 模型图 .....	7
6.3 XML Schema 和名称空间 .....	7
6.4 数据字典 .....	7
7 要求 .....	7
7.1 生态科学数据对元数据的要求 .....	7
7.2 元数据包 .....	8
7.3 元数据<<数据类型>> .....	10
7.4 生态科学数据核心元数据 .....	11
7.5 元数据扩展和元数据应用专规 .....	11
7.6 元数据应用专规示例 .....	11
7.7 元数据实施 .....	11
8 元数据模式 .....	11
8.1 元数据包 .....	11
8.2 元数据<<数据类型>> .....	19
附录 A(规范性附录) 数据字典 .....	22
附录 B(规范性附录) 元数据扩展和元数据应用专规 .....	86
附录 C(规范性附录) 抽象测试套件 .....	89
附录 D(规范性附录) 数据集全集元数据应用专规 .....	91
附录 E(规范性附录) 核心元数据 .....	188
附录 F(资料性附录) 元数据应用专规示例 .....	189
附录 G(资料性附录) 元数据应用专规制定方法 .....	207
附录 H(资料性附录) 元数据实施 .....	209
参考文献 .....	210

# 生态科学数据元数据

## 1 范围

本标准定义了一个在数据集层次上描述生态科学数据的概念模型,规定了构成该模型的一组必选的、条件必选的和可选的元数据子集、元数据实体和元数据元素,提供了对该模型的形式化描述。本标准还规定了各种需求层次的元数据应用所需要的最小元数据元素集(简称为核心元数据),以及规定了为满足特殊需求对元数据进行扩展和制定元数据应用专规的规则和方法。

本标准中的生态科学数据包括生态观测、调查、试验等生态科学研究活动中所产生的原始基本数据,也包括根据科研需求对原始基本数据进行加工整理而形成的数据,既可以是数字化的,也可以是非数字化的。

本标准适用于生态科学数据集和数据集系列的编目和描述、数据集和数据集系列的组织管理,也可用于数据交换中心的数据服务。本标准期望为生态科学数据生产者、数据管理者、数据管理系统分析和设计人员以及与数据生产、组织管理、服务或使用有关的其他人员所使用,以便理解生态科学数据元数据标准化的基本原理和总体要求。

尽管本标准主要用于生态科学数据集,但其原理可以扩展到其他领域的科学数据资源。

通过定义一个通用的概念模型,本标准保证各元数据实施之间有较高程度的语义互操作性,但是,本标准并没有规定元数据在一个信息系统中具体如何实施。

## 2 规范性应用文件

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。

GB/T 2659 世界各国和地区名称代码(GB/T 2659—2000, eqv ISO 3166-1:1997)

GB 3102.1—1993 空间和时间的量和单位(eqv ISO 31-1:1992)

GB/T 4880—1991 语种名称代码(eqv ISO 639:1988)

GB/T 7408—1994 数据元和交换格式 信息交换 日期和时间表示法(eqv ISO 8601:1988)

GB/T 7714—2005 文后参考文献著录规则(ISO 690:1987, Documentation—Bibliographic references—Content, form and structure; ISO 690-2:1997, Information and documentation—Bibliographic references—Part 2: Electronic documents or parts thereof, NEQ)

GB/T 18391.3—2001 信息技术 数据元的规范与标准化 第3部分:数据元的基本属性(idt ISO/IEC 11179-3:1994)

GB/T 19710—2005 地理信息 元数据(ISO 19115:2003, MOD)

ISO/IEC 19501:2005 信息技术—开放的分布式处理—统一建模语言(UML) Version 1.4.2  
W3C Recommendation XML Schema 1.1:2004

## 3 术语和定义

下列术语和定义适用于本标准。

### 3.1

**列表型数据 tabular data**

以表格或类表格形式存在的数据。

3.2

**栅格数据 raster data**

按栅格单元的行和列排列的、有不同灰度值或颜色的阵列数据。

3.3

**矢量数据 vector data**

以坐标串表示的空间点、线、面等几何数据及与其相联系的有关属性数据的总称。

3.4

**数据集 dataset**

可以识别的数据集合。

注1：通过对诸如覆盖范围或者属性类型的限制，数据集在物理上可以是更大数据集中的一个小数据组。

注2：改写 GB/T 19710—2005，定义 4.2。

3.5

**数据集系列 dataset series**

符合相同产品规范的数据集的集合。

[GB/T 19710—2005，定义 4.3]

3.6

**数据实体 data entity**

包含数据内容的一个逻辑的或物理的存储单元。

例如，文本文件、关系数据库数据表、电子表格、GIS 图层等。

注：一个数据集可能由一个或多个数据实体组成。

3.7

**元数据 metadata**

定义和描述其他数据的数据。

[GB/T 18391.1，定义 3.41]

3.8

**元数据实例 metadata instance**

与某一元数据模式相符的某一具体数据资源的元数据。

3.9

**元数据元素 metadata element**

元数据的基本单元。

[GB/T 19710—2005，定义 4.6]

注1：与 UML 术语中的属性同义。

注2：元数据元素在元数据实体中是唯一的。

3.10

**元数据实体 metadata entity**

一组说明数据相同特性的元数据元素。

[GB/T 19710—2005，定义 4.7]

注1：可以包含一个或多个元数据实体。

注2：与 UML 术语中的类同义。

3.11

**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。

[GB/T 19710—2005，定义 4.8]

注：与 UML 术语中的包同义。

## 3.12

**元数据注册系统 metadata registry**

登记和发布元数据的语义、结构、交换格式等的权威信息的可公开访问的系统。

## 3.13

**元数据应用专规 metadata application profile**

为特定应用定义的元数据元素、规则和指南的一个集合,这些元数据元素来自一个或多个已有的元数据标准,还可以包括用户自定义的元数据元素。

## 3.14

**编码体系 encoding scheme**

提供语境信息或句法分析规则,以帮助对元数据元素的值进行解释的受控词表或者格式规范。

注:编码体系包括词汇编码体系和句法编码体系。

## 3.15

**词汇编码体系 vocabulary encoding scheme**

作为元数据元素值域的受控词表。

例如,汉语主题词表、世界各国和地区名称代码表等。

## 3.16

**句法编码体系 syntax encoding scheme**

元数据元素的值在格式上遵从的正式表示法。

例如,“2000-01-01”作为日期的标准表示。

## 3.17

**名称空间 namespace**

名称的集合,以统一资源标识符(URI)标识。

注:作为定义名称的语境,名称空间是解决不同来源的词汇之间的名称冲突问题的一种机制。

## 3.18

**数据类型 data type**

有效值域和允许对该值域内的值进行的合法操作的规定。

例如,整型、实型、布尔型、字符串、日期以及本标准中自定义的数据类型等。

注1:数据类型包括基本预定义类型和用户定义的类型。

注2:数据类型用术语标识,例如整型。

注3:改写 GB/T 19710—2005,定义 4.1。

## 3.19

**值域 value domain**

允许值的集合。

## 3.20

**概念模式 conceptual schema**

概念模型的形式化描述。

## 3.21

**概念模型 conceptual model**

定义一个论域的概念上的模型。

## 3.22

**论域 universe of discourse**

一场讨论或辩论中所涉及的所有实体构成的范畴。



3.23

**模式 schema**  
模型的形式化描述。

3.24

**模型 model**  
对论域某些方面的抽象。

3.25

**数据模型 data model**  
以反映信息结构的某种方式对数据组织的描述。  
注：数据模型指明数据的属性、结构和内在关系。

3.26

**数据交换中心 clearinghouse**  
收集和发布数据资源的元数据,提供对分布式数据资源的统一查询和定位的一站式服务系统。

3.27

**资源 resource**  
具有身份标识的任何对象。  
例如,数据集、服务、人员或机构等。

4 符号和缩略语

4.1 缩略语

HTML	超文本置标语言(Hypertext Markup Language)
MAP	元数据应用专规(Metadata Application Profile)
MDR	元数据注册系统(Metadata Registry)
SGML	标准通用置标语言(Standard Generalized Markup Language)
UML	统一建模语言(Unified Modeling Language)
URI	统一资源标识符(Uniform Resource Identifier)
XML	可扩展置标语言(Extensible Markup Language)

4.2 UML 模型符号

本标准采用统一建模语言(UML)描述元数据子集、元数据实体和元数据元素之间的关系。用UML中的包表示元数据子集,类表示元数据实体,属性表示元数据元素。本标准中使用的主要UML符号如图1所示:

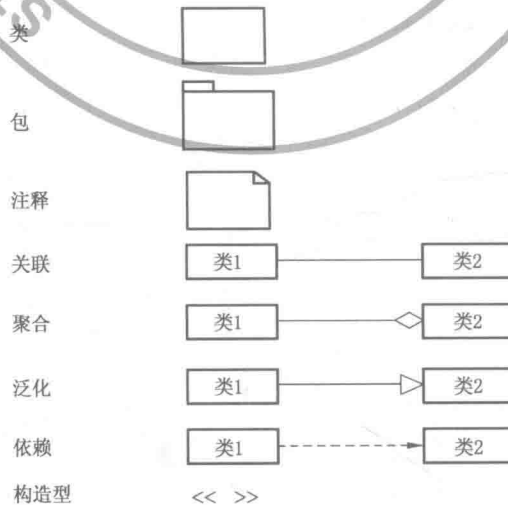


图1 UML符号



UML 图中的某一部分需要给出补充说明或明确的解释时,注释会很有用。可以把注释看成是图形化的黄页,补充性或解释性的文字写在表示注释的带折角矩形框中。注释和被注释的图元素之间用一条虚线“——”连接。

### 4.3 UML 模型关系

#### a) 关联

关联用于描述两个或更多类之间的一般关系。

关联的方向必须说明。如果不指明方向,则假定为双向关联。如果是单向关联,关联方向可以在线段终点用箭头来标记。

#### b) 聚合

聚合用于创建两个类之间的部分与整体的关系。在该关系中,一个类担当容器角色,另一个类担当容器的构件角色。聚集的表示法是从“部分”类画一条带空心菱形箭头的实线指向“整体”类。

#### c) 泛化

泛化表示父类(或超类)和可以替代它的子类之间的关系。父类是泛化类,而子类则定义为特化类。泛化的表示法是从子类画一条带空心三角箭头的实线指向父类。

#### d) 依赖

依赖用来表示对包(元数据子集)的理解、使用等依赖其他的包(元数据子集)。依赖的表示法是从依赖类画一条带箭头的虚线指向被依赖类。

#### e) 角色

两个类发生关联时,每个类通常在关联中都扮演着某种角色。UML 模型中可以使用“角色名称”标明目标对象到源对象的关系。在双向关联中,将提供两个角色名称。图 2 说明了在 UML 图中如何表示角色名称和基数。

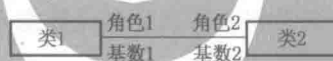


图 2 UML 角色

图 2 中,角色名称“角色 1”是类 1 对类 2 的关系,“角色 2”是类 2 对类 1 的关系,“基数 1”表示类 1 有多少个对象和类 2 的单个对象关联,“基数 2”表示类 2 有多少个对象和类 1 的单个对象关联。基数的取值可以是一个正整数  $n$  (例如,“1”、“2”、“3”),也可以是“0.. $n$ ”、“0..\*”、“ $n$ ..\*”、“ $m$ .. $n$ ”( $m$  小于  $n$ ),还可以是“ $m$ , $n$ ”( $m$  小于  $n$ )。其中, $m$  和  $n$  都是一确定的正整数;“\*”代表许多、多个;“..”在“0..\*”、“ $n$ ..\*”语境中表示“或”,例如,基数 1 为“0..\*”表示类 1 的 0 个或多个对象与类 2 的 1 个对象关联,基数 1 为“1..\*”表示类 1 的 1 个或多个对象与类 2 的 1 个对象关联;“..”在“0.. $n$ ”、“ $m$ .. $n$ ”语境中表示“到”,例如,基数 1 为“1..3”表示类 1 的 1 个、2 个或 3 个对象与类 2 的 1 个对象关联;“,”表示“或”,例如,基数 1 为“2,4”表示类 1 的 2 个或 4 个对象与类 2 的 1 个对象关联。在单向关联中,一般只标注箭头指向的类的多少个对象与箭头背向的类的一个对象关联。

### 4.4 UML 模型构造型

UML 构造型是现有 UML 概念的扩展机制。它是一种用来对其他 UML 元素进行分类(或标记)的模型元素。构造型在固有的 UML 元模型类层次结构的基础上,增强了分类机制。

本标准中使用如下构造型:

- `<<type>>`(`<<类型>>`):一个构造型类,用于定义实例(对象)的域以及作用于这些对象的操作。类型可以有属性和关联,但没有方法。
- `<<datatype>>`(`<<数据类型>>`):缺乏同一性、其操作没有副作用的一组值的描述符。数据类型包括基本预定义类型和用户定义类型。数据类型是一个具有很少操作或没有操作的类,其主要目的是容纳另一个类的抽象状态,以进行传输、编码或持久地存储。
- `<<enumeration>>`(`<<枚举>>`):一种数据类型,其实例组成一个有名称的字符值的列

表。枚举名称及其字符值均要声明。枚举意味着一个类中的可能值完全已知。

- d) <<codelist>>( <<代码表>>):用于描述更开放的枚举。它是一种灵活的枚举,用于表示可能值的一个长列表。如果该列表的元素是完全已知的,应当使用枚举;如果仅知道元素的可能值,则应使用代码表。即枚举是封闭的、不可扩展的,而代码表是可以扩展的。
- e) <<abstract>>( <<抽象>>):抽象类,该类不能被直接实例化,UML 模型中用斜体显示这个类的名称。

## 5 一致性

本标准定义的元数据实体和元数据元素在第 7 章、第 8 章和附录 A 中阐明。

用户制定的元数据应用专规按照附录 B 中的要求和规则定义和描述。

一个声称与本标准严格一致的元数据应用专规只包含本标准中定义的元数据元素、元数据实体和元数据子集,并应能通过附录 C 中所要求的测试。

一个声称与本标准一致的元数据应用专规可包含扩展的元数据元素、元数据实体和元数据子集,并应能通过附录 C 中所要求的测试。

## 6 约定

### 6.1 元数据基本属性

对于每个元数据元素和元数据实体,本标准从语义和语法两方面进行描述。具体来说,使用以下 8 个属性(基于 GB/T 18391.3—2001 定义的有关属性)进行定义:

#### a) 中文名称

赋予元数据元素或元数据实体的一个中文标记。元数据实体名称在本标准中是唯一的,元数据元素名称在元数据实体中是唯一的,通过元数据实体名称和元数据元素名称的组合,使元数据元素名称在整个标准中唯一。

#### b) 英文名称

元数据元素或元数据实体的英文名称,一般用英文全称。

#### c) 缩写名

元数据元素或元数据实体的英文缩写名称。缩写规则如下:

- 1) 缩写名在本标准范围内必须唯一。
- 2) 对存在国际或行业领域惯用英文名称缩写的,采用惯用缩写。

#### d) 定义

对元数据元素或元数据实体含义的解释,以使之与其他元数据元素或元数据实体在概念上相区别。

#### e) 数据类型

元数据元素的有效值域和允许对该值域内的值进行有效操作的规定,例如整型、实型、字符串型等。本标准也使用数据类型属性定义元数据实体、构造型和元数据关联。

#### f) 值域

元数据元素所允许的值的集合。本标准也使用值域属性定义元数据实体,用以说明该元数据实体包含的元数据元素和实体。

#### g) 约束/条件

说明一个元数据实体或元数据元素是否应当总是在元数据中选用还是有时选用。该属性可以有如下的值:必选、条件必选、可选。其中,必选指必须选用或者说被强制选用;条件必选指当满足指定的条件时必须选用;可选指可以选用也可以不选用。元数据实体和元数据元素的约束/条件属性是一个条件性概念,即,一个元数据元素或实体必选还是可选是在其“上层”元数据实体已被(强制或非强制)选用的前提下而言的。

## h) 最大出现次数

说明元数据元素或元数据实体可以有的实例的最大数目。

本标准第 8 章以 UML 静态结构图对元数据元素和元数据实体的语法特征进行说明。本标准附录 A 以数据字典的形式完整说明了每个元数据元素和元数据实体的以上 8 个属性。本标准附录 D 以 XML Schema 语言提供了对元数据元素和元数据实体的数据类型、值域以及语法特征等的定义。

## 6.2 UML 模型图

本标准使用 UML 抽象对象模型定义生态科学数据元数据的模型。在第 8 章提供的 UML 图中, 每幅图定义由相关元数据实体、元素、数据类型和代码表构成的元数据子集(UML 包), 在其他图中定义的相关元数据实体的元素予以省略, 并在元数据实体名称下圆括号内说明所在的包。6.1 规定的 8 个属性中, 反映语法特征的属性包括约束/条件和最大出现次数, 而 UML 中类的属性的多重性以及类间关系的基数可以看作是元数据元素和元数据实体约束/条件属性和最大出现次数属性的一种体现。

## 6.3 XML Schema 和名称空间

本标准还使用 XML Schema 语言对生态科学数据元数据进行定义。XML Schema 语言是万维网联盟(W3C)的一个推荐标准, 能够清晰定义元数据元素和元数据实体的数据类型、值域以及元数据实体、元数据元素之间的关系。

名称空间使得不同来源的 XML 文件可以使用相同名称的元素, 而且这些相同名称的元素可以无冲突地并入同一个 XML 文件中。由于本标准中定义的元数据实体和元数据元素可能在元数据应用专规中使用, 为本标准定义的元数据元素和元数据实体声明一个名称空间是必要的。

本标准的名称空间:

```
xmlns="http://www.gb.cn/med/2005/terms"
```

本标准中还使用了以下两个名称空间缩写, 分别针对 XML、XML Schema。

XML 的名称空间:

```
xmlns:xml="http://www.w3.org/XML/1998/namespace"
```

XML Schema 的名称空间:

```
xmlns:xs="http://www.w3.org/2001/XMLSchema"
```

## 6.4 数据字典

数据字典对每个元数据元素和元数据实体的中文名称、英文名称、缩写名、定义、数据类型、值域、约束/条件和最大出现次数等 8 个属性进行了完整说明。

在数据字典中, 角色名称用来标识元数据抽象模型关联, 并由“角色名称:”开头, 将它们与其他元数据元素区分; 对于元数据实体的值域, 用其包含的行数来说明; 也使用数据类型属性定义元数据实体、构造型和关联; 对于元数据元素和元数据实体的约束/条件属性, 用“M”表示必选, “C”表示条件必选, “O”表示可选; 对于元数据元素和元数据实体的最大出现次数, 只出现一次用“1”表示, 不固定次数的重复出现用“N”表示, 允许不为 1 的固定出现次数, 并用相应的数字(即“2”、“3”、……)表示。

## 7 要求

### 7.1 生态科学数据对元数据的要求

本标准定义描述生态科学数据所需要的元数据。元数据可以应用于独立的数据集, 也可应用于数据集系列。应当为生态科学数据集提供元数据, 也可为生态科学数据集系列提供元数据。

图 3 定义生态科学数据集、数据集系列和元数据之间的关系。它规定一个数据集应有一个或多个相关的元数据实例。数据集可以聚合为数据集系列, 也应为数据集系列提供元数据。元数据由一个或多个元数据子集(UML 包)组成, 一个元数据子集包含一个或多个元数据实体(UML 类)。

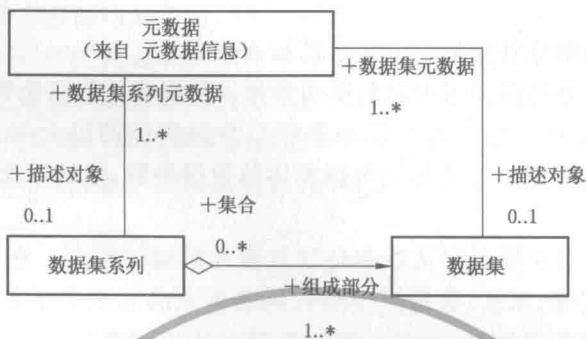


图 3 元数据应用

7.2 元数据包

7.2.1 包和元数据实体关系

本标准中,元数据子集用UML包表示。每个包包包含一个或多个元数据实体(UML类)。元数据实体包含标识离散元数据单元的元数据元素(UML类属性)。元数据实体可以与一个或多个其他元数据实体相关。元数据实体可以按需要聚合或重复以满足本标准规定的必选要求以及用户的其他要求。图4表示包的结构。第8章和附录A分别用不同的方式描述了元数据。

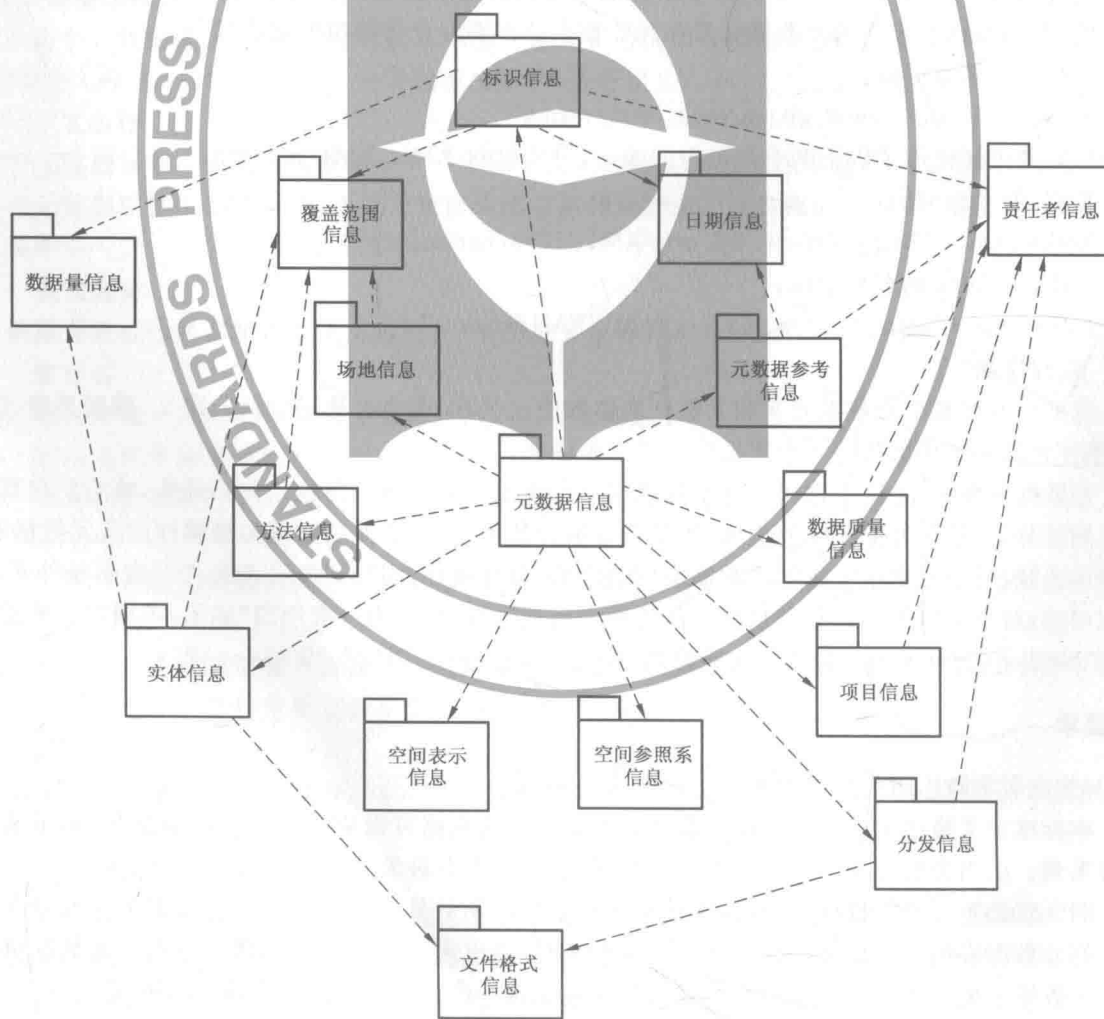


图 4 元数据包

表 1 列出了元数据的包与元数据实体之间的关系。包列在元数据包栏目下,包含在相应包中的元数据聚集实体列在元数据实体栏目下。包含在包中的元数据实体在 7.2.2 和 7.3 中进一步说明。

表 1 元数据包和元数据实体间的关系

本章条目编号	元数据包	元数据实体	UML 图和 XML Schema 定义	数据字典
7.2.2.1	元数据信息	元数据	8.1.1	A.1.1
7.2.2.2	标识信息	标识	8.1.2	A.1.2
7.2.2.3	数据质量信息	数据质量	8.1.3	A.1.3
7.2.2.4	方法信息	方法	8.1.4	A.1.4
7.2.2.5	场地信息	场地	8.1.5	A.1.5
7.2.2.6	项目信息	项目	8.1.6	A.1.6
7.2.2.7	分发信息	分发	8.1.7	A.1.7
7.2.2.8	元数据参考信息	元数据参考	8.1.8	A.1.8
7.2.2.9	实体信息	实体	8.1.9	A.1.9
7.2.2.10	空间参照系信息	空间参照系	8.1.10	A.1.10
7.2.2.11	空间表示信息	空间表示	8.1.11	A.1.11
7.3.1	责任者信息	责任者	8.2.1	A.2.1
7.3.2	覆盖范围信息	覆盖范围信息	8.2.2	A.2.2
7.3.3	日期信息	日期信息	8.2.3	A.2.3
7.3.4	数据量信息	数据量信息	8.2.4	A.2.4
7.3.5	文件格式信息	文件格式	8.2.5	A.2.5

## 7.2.2 包说明

### 7.2.2.1 元数据信息(元数据)

“元数据信息”由必选的元数据实体(UML类)“元数据”组成。后者是标识、数据质量、方法、场地、项目、分发、元数据参考、实体、空间参照系、空间表示等元数据实体的聚集。

### 7.2.2.2 标识信息(标识)

“标识信息”包含可以唯一标识数据资源的信息,包括有关数据资源的题名、摘要、目的、关键词、贡献者、状态、日期、维护、关联、限制、分类和时间空间覆盖范围等的信息。元数据实体“标识”是必选的,它包含必选、条件必选和任选的元数据元素,是关键词说明、其他贡献者、维护、关联、限制、分类以及覆盖范围等元数据实体的聚集。

### 7.2.2.3 数据质量信息(数据质量)

“数据质量信息”包含对数据资源质量的评价。元数据实体“数据质量”是任选的,在制定元数据应用专规时,根据应用需求的不同,它可以被选择也可以不被选择作为元数据应用专规的组成部分。“数据质量”是数据范围和质量报告的聚集。

### 7.2.2.4 方法信息(方法)

“方法信息”包含数据资源产生过程中使用的方法(方法是影响数据质量的重要因素)的有关信息。它包括任选的元数据实体“方法”,后者是研究方法、数据志以及质量控制元数据实体的聚集。

### 7.2.2.5 场地信息(场地)

“场地信息”包含野外观测、调查、试验等活动的发生地的有关信息。元数据实体“场地”是任选的,一般而言,它用于对数据产生的自然环境背景进行说明。

### 7.2.2.6 项目信息(项目)

“项目信息”包含作为创建数据集的研究背景的项目的有关信息。元数据实体“项目”是任选的,包

括对项目名称、动机和目标、资金来源、人员等的说明。

#### 7.2.2.7 分发信息(分发)

“分发信息”包含有关资源如何分发和获取的信息。它包括任选的元数据实体“分发”，后者是分发格式、分发订购程序、传送选项和分发联系方等的聚集。如果对外提供资源发现和访问服务，那么在制定元数据应用专规时，分发信息一般会作为元数据应用专规的组成部分。

#### 7.2.2.8 元数据参考信息(元数据参考)

“元数据参考信息”提供有关元数据实例自身(而不是元数据实例所描述的数据资源)的信息，包括元数据实例的语种、字符集、日期、联系方、依据的元数据标准等的信息。元数据实体“元数据参考”是必选的。

#### 7.2.2.9 实体信息(实体)

“实体信息”包含组成数据集的数据实体的有关信息。元数据实体“实体”是可选的，它用于对数据实体的名称、描述等基本信息以及数据实体的结构(包括逻辑结构和物理结构)进行说明。数据实体可能是列表类型的，例如关系数据库数据表、电子表格、具有固定结构的文本文件等，可能是 GIS 图层，也可能是一般图像、模型、视频文件、音频文件等。

#### 7.2.2.10 空间参照系信息(空间参照系)

“空间参照系信息”包含对数据集使用的空间参照系的说明。元数据实体“空间参照系”是可选的，是专门针对空间数据资源的，在制定针对非空间数据资源的元数据应用专规时，不需要选用该元数据实体。

#### 7.2.2.11 空间表示信息(空间表示)

“空间表示信息”包含数据集中空间信息表示方法的信息，它包括任选的元数据实体“空间表示”。“空间表示”可以特化为“栅格空间表示”和“矢量空间表示”元数据实体，“栅格空间表示”和“矢量空间表示”都包含必选的和任选的元数据元素。

### 7.3 元数据<<数据类型>>

#### 7.3.1 责任者信息(责任者)

“责任者信息”提供与资源有关的负责人或负责单位的信息。元数据实体“责任者”包括资源的负责人标识，和(或)单位，和(或)职务，还说明负责人或单位的联系信息。

构造型<<数据类型>>在 4.4 中定义。

#### 7.3.2 覆盖范围信息(覆盖范围信息)

“覆盖范围信息”提供资源所涉及空间、时间覆盖范围的信息。元数据实体“覆盖范围信息”是空间覆盖范围和时间覆盖范围的聚集。空间覆盖范围可以分为地理覆盖范围和垂向覆盖范围。地理覆盖范围又可以分为地理边界矩形、地理名称、地理多边形子类。时间覆盖范围可以分为单一时间和时间段子类。

构造型<<数据类型>>在 4.4 中定义。

#### 7.3.3 日期信息(日期信息)

“日期信息”包含与资源创建、修改、发布等生命周期中的事件相关的时间的信息。元数据实体“日期信息”是描述资源的创建时间、最近修改时间和发布时间的元数据元素的聚集。

构造型<<数据类型>>在 4.4 中定义。

#### 7.3.4 数据量信息(数据量信息)

“数据量信息”提供对数据资源大小的说明。元数据实体“数据量信息”包括数据资源的物理存储量，和(或)所包含的记录数。

构造型<<数据类型>>在 4.4 中定义。

#### 7.3.5 文件格式信息(文件格式)

“文件格式信息”提供数据对象文件格式的信息。元数据实体“文件格式”包括文件的格式名称、版



本、用来读取文件的解压缩算法或处理说明等备注信息。

构造型<<数据类型>>在 4.4 中定义。

#### 7.4 生态科学数据核心元数据

本标准定义完整的元数据元素集;通常在一个元数据应用专规中,并不需要应用全部这些元数据元素,而仅仅应用其中的一部分。然而,对任何类型的生态科学数据集而言,不管其描述需求层次如何,有一些元素都是需要和适用的。附录 F 提供了各种应用范围和需求层次都需要的最小元数据元素集(包括必选的和推荐任选的元素),称之为生态科学数据核心元数据。与本标准一致的每个元数据应用专规都应包含该核心元数据。

#### 7.5 元数据扩展和元数据应用专规

本标准第 8 章和附录 A 提供了标准元数据和关联结构,可用于多种多样的数据集。然而,资源和需求的多样性意味着通用元数据可能适应不了所有的应用,或者说不能恰如其分地满足特定的应用。这就需要根据应用需求情况对本标准中定义的元数据元素和元数据实体进行选择,并有可能需要扩充元数据元素和元数据实体。附录 B 提供了定义和应用扩充元数据元素和元数据实体的规则,以及制定元数据应用专规的规则,以更好地满足特殊用户的需求。附录 C 定义了声称与本标准一致的元数据应用专规所必需通过的测试。附录 G 从方法上为制定元数据应用专规提供指导。

#### 7.6 元数据应用专规示例

附录 F 提供中国科学院科学数据库项目根据应用需求制定的一个元数据应用专规作为示例,以帮助用户更好地理解和使用本标准。

#### 7.7 元数据实施

附录 H 提供了实施和管理元数据的一般方法和概念,以达到查询、检索、交换和表示元数据的目的。

### 8 元数据模式

#### 8.1 元数据包

##### 8.1.1 元数据信息(见图 5)

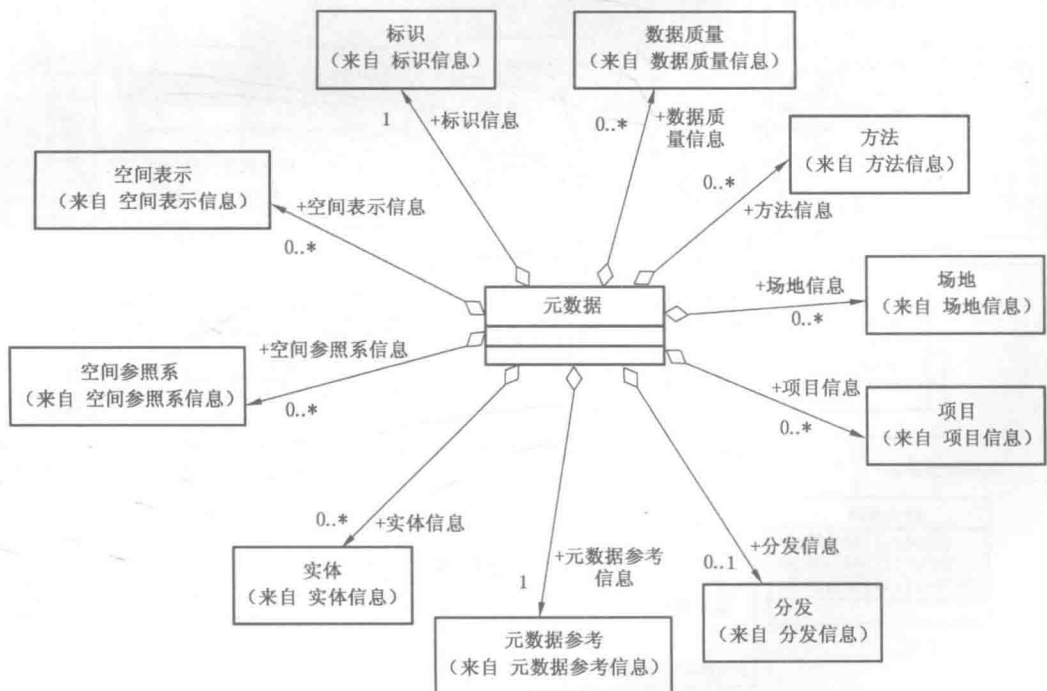


图 5 元数据信息



8.1.2 标识信息(见图6)

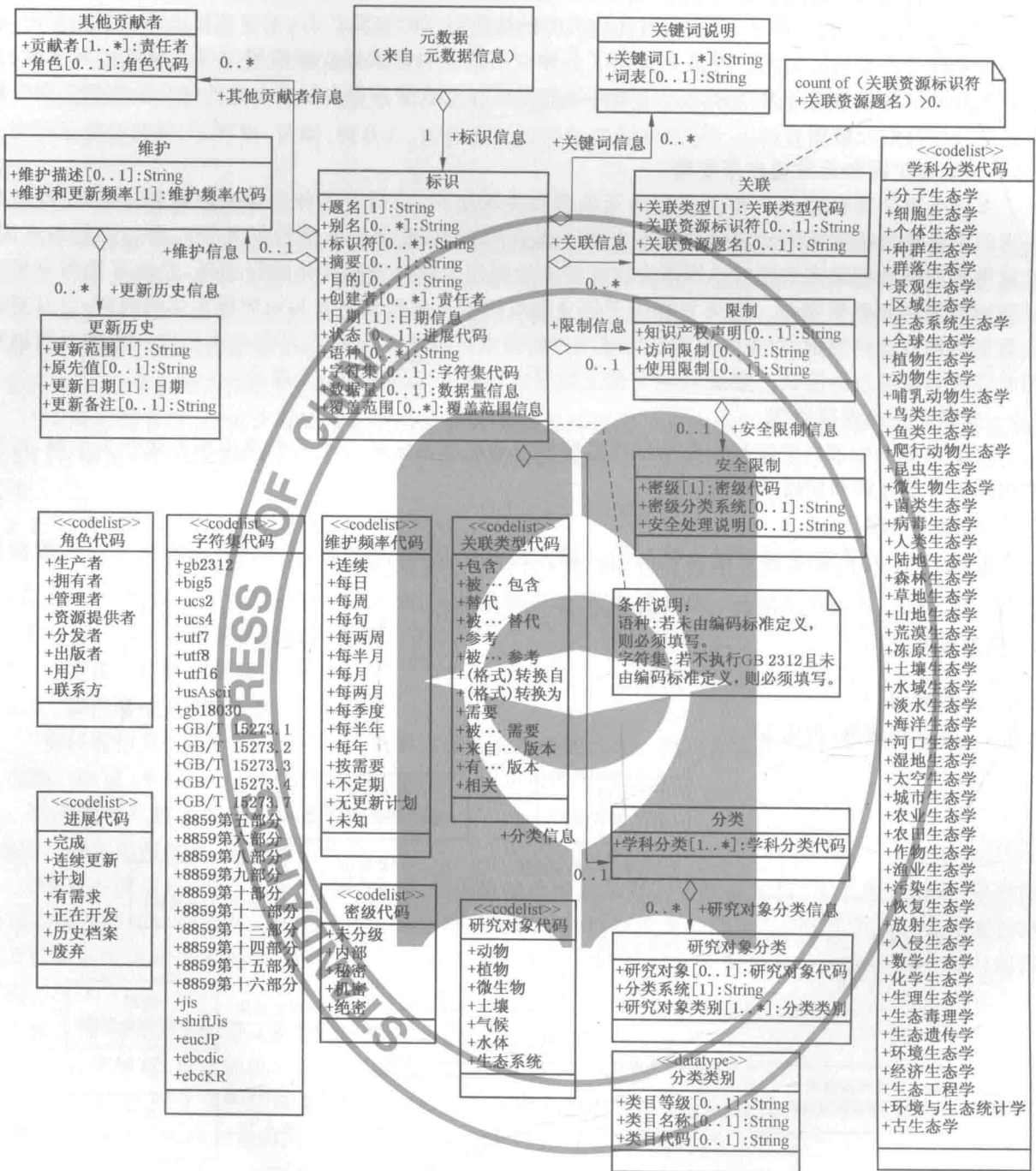


图6 标识信息

8.1.3 数据质量信息(见图7)

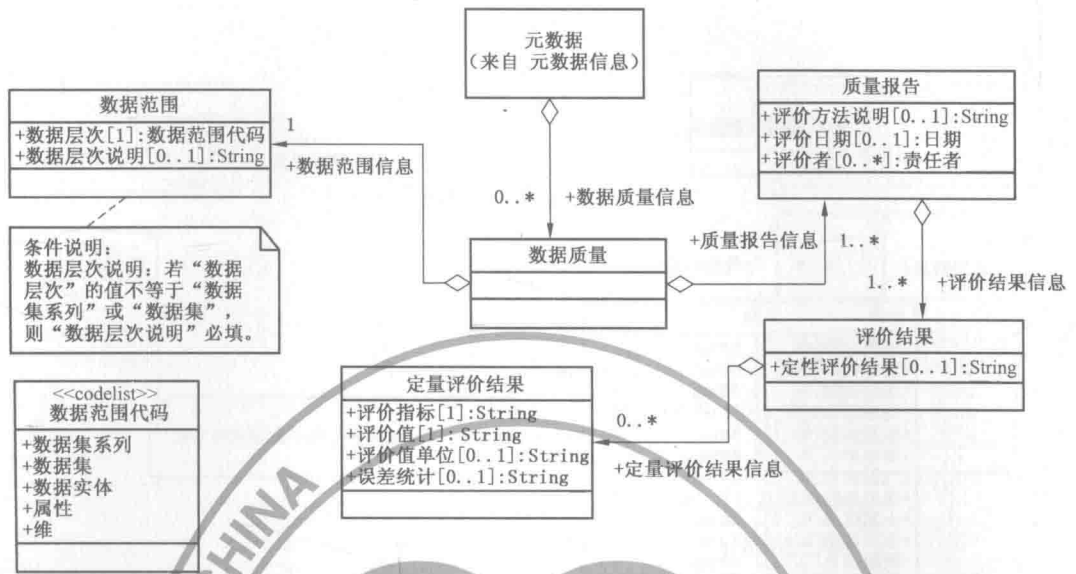


图7 数据质量信息

8.1.4 方法信息(见图8)



图8 方法信息