

ICS 35.240  
A 24



# 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

## 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

## 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。

# 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级别的语义互操作性,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。