

附件 3

# 《环境信息元数据规范（征求意见稿）》

## 编 制 说 明

《环境信息元数据规范》编制组

二〇一七年九月

项目名称	环境信息元数据规范
项目编制单位	环境保护部信息中心、环境保护部环境与经济政策研究中心
编制组主要成员	王建生、郝千婷、邢晶晶、张波、刘定、朱琦、韦正峥
环境保护部环境标准研究所技术管理负责人	李琴、王海燕
环境保护部办公厅信息办、科技标准司 环境健康管理处项目负责人	詹志明、宛悦

# 目 录

1 项目背景.....	1
2 标准修订的必要性.....	1
3 国内外研究现状.....	2
4 编制原则.....	3
5 主要技术内容说明.....	4
6 对实施本标准的建议.....	12

# 1 项目背景

## 1.1 任务由来

随着环保信息化建设的快速发展，为满足环境信息标准化工作的需要，环境保护部决定组织开展《环境信息元数据规范》（HJ 720—2014）修订工作，修订任务由环境保护部信息中心（以下称信息中心）、环境保护部环境与经济政策研究中心（以下称政研中心）承担。

## 1.2 工作过程

2017年4月：环境保护部办公厅、科技标准司向信息中心、政研中心下达了标准修订任务，确定了标准起草组的成员、组织形式、工作机制、工作计划、时间安排和工作任务分工。

2017年4月至7月：环境保护部科技标准司组织召开了《环境信息元数据规范》的开题论证会，明确了标准编制原则、方法、技术路线和标准草案的基本框架。会后，标准编制组针对《环境信息元数据规范》（HJ 720—2014）标准的应用情况展开调研，提出本轮标准修订元数据规范覆盖的范围，并形成标准草案。

2017年8月：标准编制组邀请相关信息标准专家对标准草案进行了论证，重点对标准中相关术语表达、附录修改等问题进行了讨论，标准编制组在充分吸取专家意见后形成了《环境信息元数据规范》（征求意见稿）及其编制说明。

# 2 标准修订的必要性

环境数据和信息的标准化问题已经成为生态环境大数据建设的瓶颈之一。目前，各级环境保护部门都在通过信息化建设来支持业务工作，由于环境信息标准建设滞后，特别是缺乏统一的数据字典和数据集标准对信息系统建设和数据采集进行规范，导致不同部门、不同地区采集的数据内容、定义、格式、表达不一致，极大限制了数据共享使用，甚至因为数据标准不一致，各部门来源数据无法实现整合分析，从而出现拥有大量数据而无数据可用的尴尬局面。随着信息化时代的到来，各业务部门都在采集大量数据，但数据采集过程中，经常发生需要的核心数据没有采集，而不需要的数据大量冗余的情况，如何更有针对性、更经济、更有效的采集到核心数据已经成为环保工作信息化和信息系统建设不得不解决的问题。

信息化建设，信息标准先行，这是信息化建设的基本共识。元数据是环境信息的规范化描述。这种规范化描述可以准确和完备地说明环境信息的各项特征。通过元数据，能够使用户对环境信息的定义、取值、格式、质量、处理方法和获取方法等各方面细节有更深入、更细致的了解，从而实现数据语义的统一。如何对各类环境信息元数据的功能、设计方法、表示方式、语义语法规则、属性、限制条件等进行规范，为元数据标准的建立提供需要遵照的规则和方法，是需要首先解决的问题，也就是说要首先建立制定数据标准时应该遵守的标准，是“建立数据标准的标准”。本标准可为环境信息的广泛共享和大数据应用奠定基础。

现行《环境信息元数据规范》（HJ 720—2014）发布后，在环境信息资源目录建设工作中发挥了核心作用，在一定程度上推动了环境信息资源共享。但是现行标准只对数据集的元数据进行了规范，对数据元的标准化没有进行规范，只能规范环境信息资源目录的标准化工作，未从源头解决数据采集标准不一致的问题。为此，有必要对现行《环境信息元数据规范》（HJ 720—2014）进行修订，扩展元数据规范对象，由现在的服务于建立信息资源目录，扩展到能

够保障从数据采集源头实现标准化管理，从而为今后环境信息的全面共享奠定基础。

### 3 国内外研究现状

#### 3.1 国外研究现状

许多发达国家和国际科学组织都非常重视数据的采集、管理、分发与共享工作。美国、俄罗斯、英国、法国、加拿大、日本、新加坡等国家都颁布了有关信息共享的法规。美国从1990年开始制定了一系列的数据共享政策和数据标准与规范，将分散的各部门、单位和科学家个人行为纳入到国家总体行为之中。目前信息标准的研发趋势是除了建立专业分类及编码以外，还要强调信息标准系统的整体规划，以及建立有关网络共享和网络管理方面的指标和分类，尽可能满足面向管理的分布式、实用性的数据资源网络共享的要求。国际标准化组织（ISO）在国际化标准活动中占有主导地位。从成立时起，ISO已经发表了19500多个国际标准，涵盖信息技术和制造业的几乎所有方面。关于数据元标准化的ISO/IEC 11179系列标准已从1994—2000年的第一版更新到2003—2005年的第二版。第二版ISO/IEC 11179将第一版的名称“数据元的规范与标准化”更改为“元数据注册系统”，明确了数据元的注册程序就是数据元提取和标准化的过程。元数据注册系统是支持注册功能的一个元数据数据库，用于管理数据的语义，实现对数据的标准描述和一致理解，为设计、协调、标准化、应用、重用以及交换数据奠定基础。

在环保领域，发达国家环境管理部门都非常重视环境信息标准的的管理和维护工作，发布了一系列环境信息相关标准。美国国家环境保护局（EPA）负责研究和制定各类环境计划的国家标准，提供各类有关环境信息数据集、元数据、地理信息目录的注册和查询。EPA开发的注册系统（System of Registries，简称SoR）经过多年发展，目前已包含EPA应用、模型和数据库的注册（Registry of EPA Applications, Models and Databases），环境数据集网关（Environmental Dataset Gateway），数据元注册服务（Data Element Registry Services），可重用的组件服务（Reusable Component Services），术语服务（Terminology Services），设施注册服务（Facility Registry Services），物质注册服务（Substance Registry Services）七部分。其中，数据元注册服务支持EPA及其合作系统中环境数据的管理和利用。注册服务促进了数据的一致性，并更方便用户在EPA范围内的发现和访问数据，最终促进了数据的共享使用。该服务包括自动查询和下载关键元数据，允许存在于不同机构数据系统中的字段在名称、定义、含义等方面进行展示和比较。

澳大利亚卫生与福利研究院开发了元数据在线注册系统（Metadata Online Registry，METeOR），用于国家卫生、住房和社区服务统计数据与信息的元数据标准注册。METeOR可以在线为用户提供国家授权批准的数据标准，帮助用户以存在的已经授权的数据标准创建新的数据标准，旨在为用户提供综合性的支持和帮助。METeOR在技术角度上以元数据注册系统运行，承担存储、管理和分发元数据的功能。通过METeOR，用户可以查看和下载超过2600条的数据标准。这些标准可以帮助用户避免重复创建相同或类似的数据标准，创建基于国家标准的信息系统，获得可比性好的数据。

英国电子政务元数据标准（E-GMS，e-Government Metadata Standard）制定了元素、元素精炼、编码规则，以供政府人员为他们的信息资源创建元数据或设计检索系统时使用。E-GMS的第一个版本是由简单的都柏林核心（DC）元素构成。E-GMS定义了政府公共部门电子政务元数据的结构和规则，政府部门在创建其信息资源的元数据或设计信息系统的检索系统时都

要遵循该标准。E-GMS 用于实现数据的互操作，便于公众找到所需要的政府信息和服务。

## 3.2 国内研究情况

我国数据元标准化的研究工作始于 70 年代末期，虽然起步较晚，但发展迅速，取得了不少成果。我国还专门成立了全国计算机与信息处理标准化技术委员会、数据元表示分技术委员会，组织制订了数据元的元数据标准，在全国范围内普遍采用。进入 21 世纪以后，随着信息技术的快速发展，我国相关部门开始认识到数据标准化工作对实现信息广泛共享的重要性。随后，各部委均加强了数据标准建设工作。2001 年，国家启动了“科学数据共享工程”，2004 年底，国家环境保护总局组织启动“环境科学数据库建设与共享”项目，提出“环境信息元数据规范”，供项目内部使用。科技部组织专家经过研究分析美国联邦地理数据委员会元数据标准 (FGDC)、国际标准化组织 (ISO) 元数据标准等，结合我国实际情况，制定出了《NREDIS 空间元数据内容标准》。国家地理空间信息协调委员会组织国家信息中心等单位在《NREDIS 空间元数据内容标准》基础上，参考 ISO19115 及 FGDC，起草了《国家空间信息基础设施元数据内容标准》。国土资源部信息中心牵头组织国家基础地理信息中心等单位制定了《国土资源信息核心元数据标准》。国家气象局参照国际气象组织的统一元数据标准制定了《气象元数据格式标准》，该标准已实际应用与国家气象局的元数据建库和元数据发布服务系统的建设中。在卫生系统，原卫生部 2009 年发布了《卫生信息数据集分类与编码规则》(WST 306—2009)、《卫生信息数据集元数据规范》(WST 305—2009)、《卫生信息数据模式描述指南》(WST 304—2009)、《卫生信息数据元标准化规则》(WST 303—2009)、《健康档案基本架构与数据标准(试行)》、《基于健康档案的区域卫生信息平台建设指南(试行)》共 6 项信息类指导性标准。2010 年卫生部卫生信息标准委员会开始编制国家卫生数据字典，建立了国家卫生数据字典和元数据管理信息系统，并已经开始向公众提供服务。

环境保护部为了支持业务司局、各单位环境基础数据的利用、共享，也制定出了一系列国家环境数据标准以及相关数据传输标准。环境保护部 2008 年发布了《环境信息术语》(HJ/T 416—2007)、《环境信息分类与代码》(HJ/T 417—2007)、《环境信息系统集成技术规范》(HJ/T 418—2007)、《环境数据库设计与运行管理规范》(HJ/T 419—2007) 4 项信息类指导性标准，2010 年发布了《环境信息化标准指南》(HJ/T 511—2009)，2014 年发布了《环境信息元数据规范》(HJ 720—2014) 等 12 项标准。截止目前，环保系统共发布环境信息类标准 53 项。然而，目前的环境信息标准在数据元标准、基本数据集规范等直接服务于数据共享的标准方面依然滞后。

2012—2015 年，环境保护公益行业专项支持环境保护部环境与经济政策研究中心等单位开展了环境与健康风险评估基本数据集与信息共享关键技术研究，项目组对环境相关信息标准体系开展了系统性研究，分析了目前环境信息共享方面存在的主要问题，提出了建立环境信息元数据标准的整体框架，建立了国家环境与健康数据字典及元数据注册系统，并开始编制环境与健康样品采集基本数据等数据集规范指导实际业务工作中的信息采集，为《环境信息元数据规范》(HJ 720—2014) 标准的修订提供了可供借鉴的经验。

## 4 编制原则

### 4.1 充分考虑环境信息资源及当前环境信息化发展现状

环境信息资源是各级环保部门在履职过程中采集、存储、整理、加工、汇总和统计的各

种环境数据。每年各级环保部门产生和积累大量监测、监控、统计等数据，具有数据量大、数据类型多样、分布广泛等特点。另外，随着生态环境大数据平台建设，目前环保部门已经运行了超过 100 个信息系统。标准编制过程中，必须要充分考虑现有信息资源和信息系统的情况。

## 4.2 满足环境保护工作业务需求

元数据规范是通用型的技术标准，因此需要考虑不同环境保护专业领域的需要。在标准编制过程中，要针对不同业务场景进行分析，不但能够保障本规范能够满足各项环境保护业务信息化发展的个性化需要，同时也要保障今后数据标准化的实现，为数据广泛的共享、交换奠定基础。

## 4.3 标准兼容原则

环境信息资源作为政府信息资源的一部分，在制定环境信息资源元数据标准过程中，需从长远考虑，尽量和已有的国家、行业以及相关标准兼容，例如《信息技术元数据注册系统》（GB/T 18391—2009）等。

# 5 主要技术内容说明

## 5.1 适用范围

本标准规定了环境信息的元数据标准框架，并对对象类、特性、分类方案、值域、数据元概念、数据元、数据集、数据集规范、指标、数据质量声明、术语共 11 个管理项的元数据进行了规范，适用于环境信息元数据标准的管理、环境信息资源目录建设、国家最小数据集的开发、数据元字典的编制以及环境信息元数据注册系统的设计与开发。

## 5.2 编制依据

环境信息元数据规范在国内现有元数据标准规范的基础上，结合国际上现有的相关标准，制定出符合环境信息标准化描述需求的元数据规则。针对数据元的元数据，本标准主要以《信息技术元数据注册系统》（GB/T 18391—2009）标准为基本依据进行编制；针对数据集的元数据，本标准主要以《政务信息资源目录体系 第 3 部分 核心元数据（GB/T 21063.3—2007）》为基本依据进行编制。

## 5.3 层次框架

本标准由 7 个部分组成，包括：

- (1) 适用范围
- (2) 规范性引用文件
- (3) 术语和定义

(4) 环境信息元数据标准框架：规范了 11 个元数据管理项之间的关系，确定了环境数据标准管理的整体框架。

(5) 符号与约定：介绍了本标准定义环境信息元数据时采用的描述方式、语法规则以及 UML 表达相关要求。

- (6) 元数据项的元数据标准：规定了 11 个元数据管理项的元数据描述规则，包括通用属

性和特殊属性，以及这些属性的具体定义和规则。

(7) 附录：为便于标准用户理解，对标准中涉及到的分类取值进行了举例说明。

## 5.4 修订内容

### 5.4.1 扩展了管理对象

本次修订由现行 HJ 720—2014 标准中仅对数据集的元数据进行规范扩展为对对象类、特性、分类方案、值域、数据元概念、数据元、数据集、数据集规范、指标、数据质量声明、术语共 11 个管理项进行规范。本次修订管理对象的扩展是为了满足从数据采集源头进行数据标准化的需要设置的，涵盖了“元数据规范注册管理”和“数据应用”两个方面。具体管理项如下：

(1) 对象类。对象类是现实世界中的想法、抽象概念或事物的集合，有清楚的边界和含义，并且特性和其行为遵循同样的规则而能够加以标识。对象类的例子有人、家庭、动物、植物等。图 1 是对象类和相关元数据项之间的关系结构图。

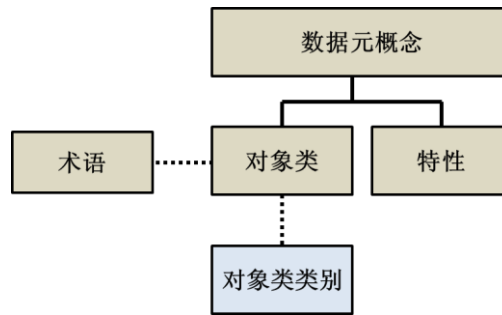


图 1 对象类与其他元数据项关系结构图

对象类类别是对象类按照特定的意义分类构成的组和亚组，主要用于在元数据注册系统中帮助用户浏览和查找相关的对象类，例如在元数据注册系统中可以包括以下八个对象类类别：

- ① 环境污染源
- ② 环境污染物（化学物质，生物物质，声，光，辐射等）
- ③ 环境（社会环境，水环境，土壤环境，大气环境等）
- ④ 人
- ⑤ 其他生物（动物，植物，微生物等）
- ⑥ 事件（环境事故，环境污染健康损害事件等）
- ⑦ 活动（监测，干预，调查，评估等）
- ⑧ 资源（设施，仪器设备，方法，标准等）

(2) 特性。特性是一个对象类的所有成员所共有的某种特征。它用来区别和描述对象类，是对象类的特征，它们构成对象类的内涵。例如血铅水平、身高等都是儿童这个对象类的特性，铅质量浓度、镉质量浓度都是土壤这个对象类的特性。特性有可能需要用多个词组加以描述。图 2 是特性和相关元数据项之间的关系结构图。

特性依据其特征可分为不同的亚组，主要用于在元数据注册系统中帮助用户浏览和查找相关的特性。例如在元数据注册系统中可以包括以下特性亚组：

- ① 位置特性（如经度，纬度，高度等）



- ② 时间特性（事故发生时间、生产时间、发病时间等）
- ③ 生产经营活动相关特性（生产时间、SO<sub>2</sub> 排放量等）
- ④ 物理特性（如浓度，体积，颜色，速度，密度，温度等）
- ⑤ 化学特性（如 COD，腐蚀性，酸碱度）
- ⑥ 毒理学特性（如 LD<sub>50</sub>，，致癌，致畸，致突变等）
- ⑦ 社会人口文化特性（年龄、性别、职业等）
- ⑧ 生理生化特性（收缩压、血脂水平、血红蛋白含量等）
- ⑨ 行为模式特性（饮水量、室外停留时间等）
- ⑩ 内暴露相关特性（血铅浓度、尿镉浓度等）
- ⑪ 健康效应特性（如功能异常，致残，疾病，死亡等）
- ⑫ 测量特性（如采样，分析，仪器，设备，试剂等）
- ⑬ 资源特性（如价值，数量，可用状态，隶属关系等）
- ⑭ 标识特性（如名称，编码等）

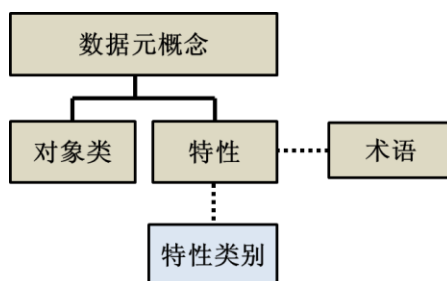


图 2 特性与其他元数据项关系结构图

(3) 分类方案。值域是基于分类方案创建的。分类方案是由国家或国际机构认可的，用来对数据进行分类的官方语言。比如国家行政区划代码、污染源分类代码、国际疾病标准分类代码 ICD-10、国际标准职业分类等。图 3 是分类方案和相关元数据项之间的关系结构图。

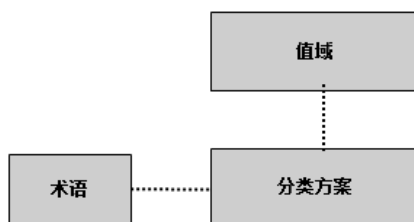


图 3 分类方案关系结构图

(4) 值域。表示可包括值域、数据类型、表示类（可选）和计量单位四部分，其中任何一部分发生变化都成为不同的表示。值域是数据元允许值的集合，例如土壤中镉的含量这一数据元的值域是用非负实数集（以 mg/kg 为单位）作为它的允许值集合；数据类型是表达数据元允许值的不同值的集合。例如被企业名称的数据类型是“字符”；表示类是数据元值域类别，例如污染源代码这一值域的表示类是“代码”；计量单位是用于计量相关值的实际单位，例如土壤中镉浓度测量值的计量单位是“mg/kg”。图 4 是值域和相关元数据项之间的关系结构图。

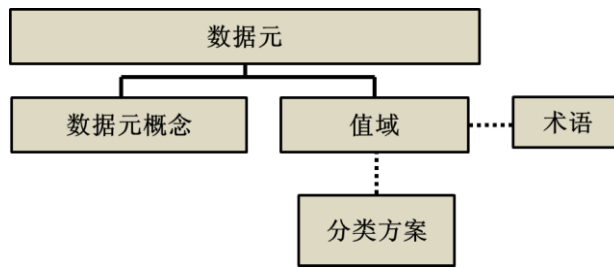


图 4 值域关系结构图

(5) 数据元概念。一个数据元概念是由对象类和特性两部分组成，是能以一个数据元形式表示的概念，其描述与任何具体的表示无关。一个数据元概念可对应多个数据元。当一个数据元概念与一个表示联系在一起时，就产生了一个数据元。一个数据元概念可以与不同的表示关联产生不同的数据元。例如，环境空气中多环芳烃浓度这个数据元概念，可以用不同的度量衡单位进行表示从而形成不同的两个数据元。图 5 是数据元概念和相关元数据项之间的关系结构图。

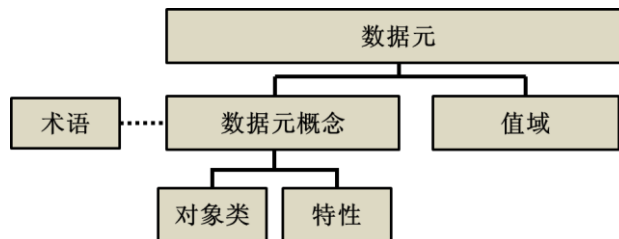


图 5 数据元概念与其他元数据项关系结构图

(6) 数据元。数据元是在特定的语义环境中被认为是不可再分的最小数据单元。数据元不是一般的数据概念，它由三部分组成：对象类、特征类和表示类（值域）三部分组成。

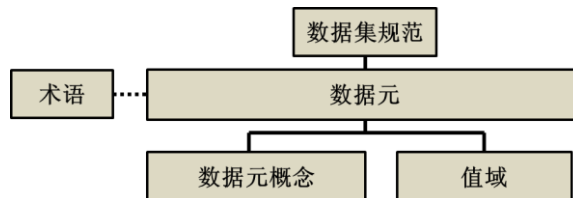


图 6 数据元与其他元数据项关系结构图

(7) 数据集规范。数据集规范明确了一项具体业务活动所必须采集的数据，提供了与数据集采集相关的数据项目的统一定义，目的是确保数据采集的标准化。其中在国家层面上要求强制收集的最少的数据元组合，就称为国家最小数据集（National Minimum Data Set, NMDS），也被称为国家基本数据集。图 7 是数据集规范和相关元数据项之间的关系结构图。

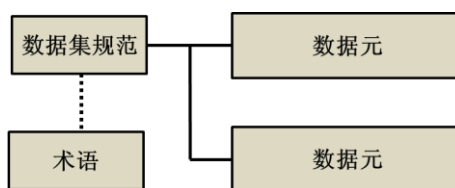


图 7 数据集规范与其他元数据项关系结构图

(8) 术语。术语是对专用概念或容易引起歧义概念的详细解释。例如，要收集企业特征污染物排放量数据，首先明确对“特征污染物”的定义。术语可以链接到任何其他元数据项目，以对专名词进行详细的解释。图 8 是术语和相关元数据项之间的关系结构图。

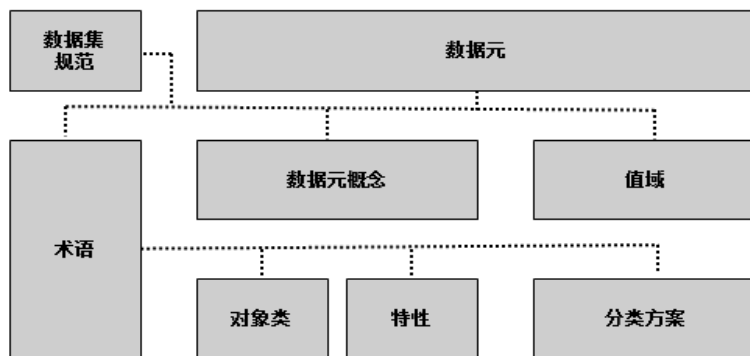


图 8 术语关系结构图

(9) 指标。指标是用来反应或评价某一事物特征的概念，包括定性指标和定量指标两种主要形式。指标的形成依赖于是否能够用于比较和评价。比如土壤重金属超标率、儿童血铅超标率等。为了服务于数据应用，本标准把指标也作为元数据项进行注册管理。

(10) 数据集。数据集是一个特定的数据资源、数据库和数据来源的引用。数据集元数据项的作用在于形成数据资源目录，为了解数据资源的内容以及后续获取数据资源实体提供信息。

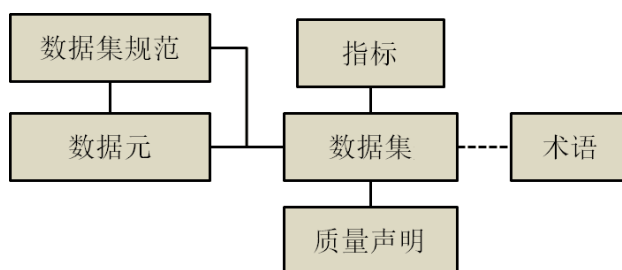


图 9 数据集与其他元数据项的关系结构图

(11) 数据质量声明。对数据质量的多维度声明，包括数据库的适用性、准确性、及时性等，目的是保证数据资源使用者能够比较全面客观的了解数据质量情况。数据质量声明对数据资源并不是强制的，但高度推荐对数据资源开展质量评价。数据质量声明可以为数据资源使用者提供更有价值的信息。

#### 5.4.2 对术语和定义进行了补充

本次修订对管理项进行了扩展，因此需要对于涉及到的术语和定义进行补充。为了与国家已经发布的相关标准保持一致，本次修订中相关术语和定义大部分采用全文引用的形式。为了便于标准用户查询术语和定义的来源，参照GB 18391标准的表达方式，在标准正文中进行了标注。术语和定义包括：

(1) 数据 (data)：信息的可再解释的形式化表示，以适用于通讯、解释或处理。原文引

用于 GB/T 5271.1—2000 中定义 01.01.02。

(2) 数据元 (data element): 由一组属性规定其定义、标识、表示和允许值的数据单元。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.8。

(3) 元数据 (metadata): 定义和描述其他数据的数据。原文引用于 GB/T 18391.1—2009 中术语和定义 3.2.16。

(4) 概念 (concept): 通过对特性的独特组合而成的知识单元。原文引用于 GB/T 15237.1—2000 中定义 3.2.1。

(5) 定义 (definition): 描述一个概念, 并区别于其他相关概念的表述。原文引用于 GB/T 15237.1—2000 中定义 3.3.1。

(6) 元数据对象 (metadata object): 由元数据标准框架定义的一个对象类型, 是需要纳入元数据规范管理的对象。GB/T 18391.1—2009 中术语 3.2.18 对元数据对象的定义为“由元模型定义的一个对象类型”, 为了便于理解, 在本规范中将元模型修改为元数据标准框架, 并进行了进一步的解释。

(7) 对象类 (object class): 可以对其界限和含义进行明确的标识, 且特性和行为遵循相同规则的观念、抽象概念或现实世界中事务的集合。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.22。

(8) 特性 (property): 一个对象类所有成员所共有的特性。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.29。

(9) 表示 (representation): 数据元的值域、数据类型的组合, 必要时也包括计量单位或字符集。原文引用于 GB/T 18391.1—2002 中术语 3.60。

(10) 表示类 (representation class): 表示类型的分类。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.35。

(11) 分类方案 (classification scheme): 根据对象具有的共同特征将对象进行排列或分组的描述性信息。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.4。

(12) 值 (value): 数据的值。原文引用于 GB/T 18391.1—2009, 术语和定义 3.3.37。

(13) 允许值 (permissible value) 在一个特定值域中允许的一个值含义的表达。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.28。

(14) 值域 (value domain) 允许值的集合。原文引用于 GB/T 18391.1—2009, 术语和定义 3.3.38。

(15) 值含义 (value meaning) 一个值的含义或语义内容。原文引用于 GB/T 18391.1—2009, 术语和定义 3.3.39。

(16) 计量单位 (unit of measure) <值域>用于计量相关值的实际单位。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.36。

(17) 数据元概念 (data element concept) 能以数据元的形式表示的概念, 其描述与任何特定表示法无关。原文引用于 GB/T 18391.1—2009 中术语和定义 3.3.9。

(18) 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。原文引用于 WS/T 305—2009 中术语和缩略语 3.1.1。

(19) 数据集规范 (dataset specification) 指针对特定业务活动利益相关人一致认同为了达到最佳业务实践而需要采集的数据元, 以及对这些数据元的标准化表达。参考澳大利亚卫生与福利研究院元数据在线注册系统中对数据集规范的解释 (A data set specification is a metadata set that is not mandated for collection but is recommended as best practice. 数据集规范是一种元数据集合, 它不是为收集数据而进行的授权, 而是为了达到最佳实践推荐收集的数据)。

### 5.4.3 增加了环境信息元数据标准框架

环境信息元数据标准框架规定了环境信息元数据管理项的内容和功能，以及这些元数据管理项之间的关系，确定了环境领域数据标准化管理工作的基本思路，为环境数据标准的建立提供了需要遵照的规则和方法。同时，在今后的元数据注册系统建设中，元数据标准框架将起到导航作用，引导元数据进行“向导式”注册。

本次标准修订中，围绕“数据标准化”和“数据应用”两个方面构建了环境信息元数据标准框架。首先从保障数据采集标准化的需要出发设置 7 种元数据对象，分别为对象类、特性、分类方案、数据元概念、值域、数据元、数据集规范。其次从数据应用角度出发，设置 3 种元数据对象，分别为指标、数据集、质量声明。同时将术语作为单独的元数据项目进行注册管理。环境信息元数据标准框架如图 10 所示。

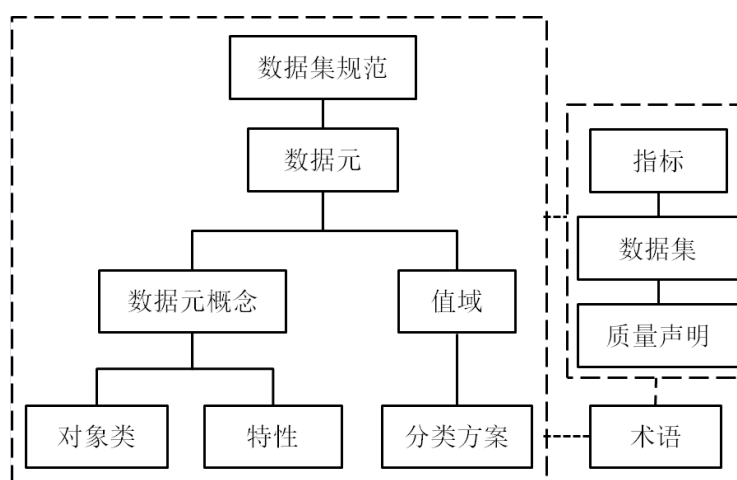


图 10 环境信息元数据标准框架

对象类和特性构成数据元概念，依据分类方案可以制定值域（虚线表示值域的制定可以不依据分类方案），数据元概念和值域构成数据元，若干数据元组成数据集规范。

指标可以由作为分子、分母的数据元计算得出，也可以从数据集中提取；数据集是可以利用的信息资源，其数据集的元数据标准是建立信息资源目录的依据；质量声明对数据集从不同维度进行质量评价，保证数据集使用者能够了解数据的质量情况。

术语是对专用概念或容易引起歧义概念的详细解释，可以链接到任何曾引用过术语的元数据项目。

### 5.4.4 调整了数据集的元数据属性

本次修订对现行HJ720—2014规定的数据集的元数据属性进行了调整（表1），包括将原来的8类属性修改为9类，新增了管理类和关系类，删除了元数据描述信息，并取消了核心元数据及属性内的层级关系。对覆盖范围类、质量类属性约束条件进行了修改，以使用户更明确对空间和时间的描述；内容属性中新增了特征数据元，以使用户了解数据集中包含的具体信息；对标识类和维护类属性中包含的子属性进行完善，以便于用户更方便的获取所需信息。本次修订数据集元数据属性是为了与目前正在实施的环境信息资源著录工作的衔接和满足实际应用中的需要，同时兼顾了整个元数据标准在管理对象扩展后属性间的一致性。

表 1 数据集的元数据属性修订前后比较

现行 HJ720—2014 中规定的属性				修订后属性		
属性种类	属性名称	子属性	约束	属性种类	属性名称	约束
标识信息	数据集名称		M	标识类	中文名称	M
	数据集发布日期		M		数据集发布日期	M
	数据集摘要		M		数据集摘要	M
	关键字		M		关键字	M
	数据集分类	类目名称	M		数据集类目名称	M
		类目编码	M		数据集类目编码	M
		分类标准	M		数据集分类标准	M
	数据量		O		数据量	O
	标识符		M		标识符	M
	数据集保管期限		O		数据集保管期限	O
	数据集提供方	数据集提供单位	M		提交机构	M
		数据集提供方电话	O			
		数据集提供方地址	O			
		数据集提供方电子邮件	O			
				英文名称	M	
				版本	M	
				元数据项类型	M	
覆盖范围信息	空间范围	描述	C	空间范围描述	C	
		地理边界矩形东边经度	M	地理边界矩形东边经度	C	
		地理边界矩形西边经度	M	地理边界矩形西边经度	C	
		地理边界矩形南边纬度	M	地理边界矩形南边纬度	C	
		地理边界矩形北边纬度	M	地理边界矩形北边纬度	C	
		垂向覆盖范围最小值	M	垂向覆盖范围最小值	C	
		垂向覆盖范围最大值	M	垂向覆盖范围最大值	C	
	时间范围	时段起始时间	M	起始时间	C	
		时段结束时间	M	结束时间	C	
		时刻	C	时刻	C	
内容信息	资源域	M	内容类	资源域	M	
				特征数据元	M	
维护信息	更新频率	M	维护类	更新频率	M	
				数据集维护机构名称	M	
				数据集维护机构电话	M	
				数据集维护机构地址	M	
				数据集维护机构电子邮件	M	
				数据集更新日期	O	

现行 HJ720—2014 中规定的属性				修订后属性		
限制信息	数据集安全限制分级		O	限制类	数据集安全限制分级	O
	数据集信息公开属性		O		数据集信息公开属性	O
数据质量信息	数据志	数据志说明	C	数据质量类	数据志说明	C
		数据源说明	C		数据源说明	C
		处理步骤	C		处理步骤	C
	数据质量报告	质量说明	C		数据质量说明	C
		数据质量评价	C		数据质量评价	C
分发信息	在线资源链接地址		O	分发类	在线资源链接地址	O
	数据分发格式		M		数据分发格式	M
	费用		O		费用	O
元数据描述信息	元数据标识符		M	关系类	相关术语	O
	元数据维护方	元数据维护方电话	O		相关指标	O
			元数据维护方地址	O	注册机构	M
	元数据更新日期		O	管理类	注册状态	M
					注册日期	M

#### 5.4.5 删除附录内容

本次修订删除了现行HJ720—2014中附录内容，主要是基于以下考虑：

- (1) 本次标准修订是对环境信息元数据标准进行整体设计，现行标准中附录中内容与新修订标准定位不一致；
- (2) 在本次对数据集的元数据属性进行调整后，附录中内容已经不适用；
- (3) 现行标准中附录中主要为数据集元数据的应用实例，这些实例在近几年的业务发展中已经有较大变化，附录内容已经与实际不符。

## 6 对实施本标准的建议

### 6.1 由专业机构负责环境信息标准维护工作

标准建设是信息化建设的基础。本标准发布后，环境保护各业务领域均需要编制基本数据集标准，以实现数据采集阶段的标准化，为数据共享奠定基础。这些工作需要专业的机构进行管理，建议环境保护部信息中心成立专门的信息标准技术管理部门，承担相应的信息标准编制的规划、组织、技术评审、标准协调等工作，从而实现环境信息标准的动态维护和可持续发展。

### 6.2 尽快完善环境信息标准体系

本标准仅为今后制定环境信息标准提供了指导，规范了各类元数据标准之间的关系，还需要在本标准的基础上尽快完善信息标准体系，才能实现数据的广泛共享。比如，针对数据集规范，应该制定数据集规范分类标准、针对数据字典的值域应该尽快完善环境信息代码标

准、针对数据质量元数据标准制定数据质量评价技术规范等。各业务司局，应该梳理各自的业务需求，针对特点的业务活动编制一系列的基本数据集标准。

### **6.3 建设环境信息元数据注册系统**

本标准发布后，将会有大量的数据元、数据集规范需要进行管理，如何实现标准的标准化管理，避免数据标准之间的矛盾是亟待解决的问题。从数据标准管理的发展趋势来看，通过建立元数据注册系统对环境信息标准进行管理是必由之路。因此，建议由环境保护部信息中心建设环境信息元数据注册系统，为广大用户提供信息标准服务，推动标准的实际应用。

---