

附件三：

**《环境数据集说明文档格式标准》（征求意见稿）
编制说明**

《环境数据集说明文档格式》编制组

二〇一〇年八月

项目名称：环境数据集说明文档格式

项目统一编号：1519

项目承担单位：环境保护部信息中心、北京市倍思电子数据库工程公司

编制组主要成员：徐富春、刘定、孙强、徐枫、宦茂盛、石雯雯

标准所技术管理负责人：李晓倩、卢延娜

标准处项目负责人：李晓弢

目 录

1	项目背景.....	1
1.1	任务来源.....	1
1.2	工作过程.....	1
2	标准制订的必要性分析.....	1
3	标准编制的依据与原则.....	2
3.1	标准编制的依据.....	2
3.2	标准编制的原则.....	4
4	标准主要技术内容.....	4
4.1	标准适用范围.....	4
4.2	标准结构框架.....	4
5	对实施本标准的建议.....	5

《环境数据集说明文档格式》编制说明

1 项目背景

1.1 任务来源

项目编号：1519

任务来源：本标准任务来源于国家环境保护标准制修订十一五计划，下达计划文件为《关于开展 2009 年度国家环境保护标准制修订项目工作的通知》（环办函[2009] 221 号），本标准由环境保护部信息中心和北京市信思电子数据库工程公司共同承担。

1.2 工作过程

1、编制任务正式启动

2009 年 12 月 14 日，环保部科技标准司在环保部六楼中厅会议室主持召开了《环境数据集说明文档格式》的开题讨论会。会议听取了专家的意见，确定了标准制定总体技术路线、标准制定实施方案和标准项目组的工作人员。

2、国内外情况调研

（1）国外情况调研

随着社会生产力的发展，信息在自然、社会、经济中的作用日益重要，世界上各个国家都将研究数据及其标准化的问题放在重要位置之上，数据集的标准化方面更是覆盖从顶层的方法论标准到具体的数据集规范。标准编制组主要调研了 ISO 19131:2007 Geographic information – Data product specifications、ISO/IEC 11179 Information Technology – Metadata registries (MDR)和美国环保局的数据标准。

（2）国内情况调研

目前，ISO 19131:2007 Geographic information – Data product specifications 尚未转化为国标，但在我国卫生和人口计生领域，均建立了各自的数据集规范。标准编制组重点调研了卫生领域的电子病历系列标准、健康档案系列标准等，以及人口计生领域的《人口和计划生育基础数据标准体系》。此外，还调研了国标 GB/T 18391《信息技术 数据元的规范与标准化》系列标准，科学数据共享工程和气象领域的有关标准规范。

3、标准征求意见稿编写工作

在前期调研的基础上，结合标准编制组以往标准制定的经验，于 2010 年 6 月形成了《环境数据集说明文档格式》的征求意见稿。

2 标准制订的必要性分析

本标准的制订是“环境信息与统计能力建设项目”建设的需要。

“环境信息与统计能力建设项目”以贯彻落实党中央、国务院关于节能减排工作部署为指导，以实现“十一五”期间重点污染物减排的目标指标为紧要任务，围绕建立与完善“科学的减排指标体系、准确的减排监测体系、严格的减排考核体系”的要求，加强数据传输、共享和应用能力、业务应用支撑能力、统计基础能力等环保信息化能力的建设，为实现“十一五”节能减排和环境保护工作目标奠定基础。

其中，减排综合数据库的建设本项目建设的主要内容，也是国家减排工作重要内容和基础环节。本项目要建立环境保护部和省局两级减排综合数据库，为减排工作提供安全、高

效的数据汇集、存储、管理、备份的基础环境。而减排数据库建设最重要的是统一各类信息资源的描述方法和原则。因此，需要建立统一的与业务相结合的数据标准，来指导数据管理和数据库建设。

《环境数据集说明文档格式》正是基于上述现实要求而提出的。数据作为“环境信息与统计能力建设项目”建设的基础与核心，数据集是该项目数据组织的主要方式，因此对数据集进行统一规范的说明和描述显得尤为重要。一方面，环境数据集的划分是否科学，将直接影响数据组织、生产、交换和共享的效率和质量；另一方面，环境数据集要进行共享，就必须建立起大家一致认可的相关标准或规范，否则会导致大量的转换工作甚至无法实现共享。为此，建立《环境数据集说明文档格式》，采用科学有效的方法对数据集进行划分，并对数据集进行规范化描述，可有效规避风险，提高数据组织和生产的效率和质量。

3 标准编制的依据与原则

3.1 标准编制的依据

在标准编制过程中，我们重点调研了国内外相关标准情况，其中 ISO19131:2007 Geographic information -- Data product specifications 地理信息数据产品规范是本标准编制的重要依据。

1、国外相关标准情况

(1) ISO19131:2007 Geographic information -- Data product specifications

ISO19131:2007 Geographic information -- Data product specifications 地理信息数据产品规范，是数据集或数据集系列的详细描述及补充说明，从而使其他方能够根据规范创建、提供和使用数据集或数据集系列。就数据产品必须或可能达到的要求而言，数据产品规范是数据产品的一个准确的技术描述。然而由于种种原因，在执行过程中可能需要作出妥协，数据产品规范只是定义了数据集应当如何，与产品数据集有关的元数据应该反映产品数据集的实际状况。

数据产品规范可能由不同的组织在不同的场合因不同的原因制定和使用。例如，规范可能用于数据采集的原始过程，也可能用于从现有数据中派生产品。生产者可能用它来规范自己的产品，用户也可能用它来陈述需求。

该国际标准的目的在于为制定数据产品规范提供实用的帮助，并与其他现有的地理信息标准保持一致，其目标是提出用于规范数据产品的诸项目的明细表。

对于数据产品规范而言，只需要规范最终数据产品的某些项而不必对产品处理过程进行详细描述。然而，若认为有必要，规范也可以包括数据生产和维护等方面的内容。

该国际标准描述了数据产品规范的内容与结构。

该国际标准规定数据产品规范应包括覆盖数据产品以下方面的主要部分：概述、规范范围、数据产品识别、数据内容和结构、参照系、数据质量、数据产品分发、元数据。数据产品规范还应包括涉及数据产品以下方面：数据获取、数据维护、图示表达和附加说明。

(2) ISO/IEC 11179 Information Technology -- Metadata registries (MDR)

国外的数据标准化工作起步较早。1987 年国际标准化组织（ISO）和国际电工委员会（IEC）在原计算机和信息处理（ISO/TC 97）、信息技术设备（IEC/TC 47B）的基础上联合成立了信息技术标准化联合技术委员会（ISO/IEC JTC1）。ISO/IEC JTC1 成立后，数据元标准化工作得到进一步加强。SC32 为 JTC1 下设的数据交换与管理分技术委员会，该分技术委员会由 4 个工作组构成，其中，SC32/WG2 工作组以数据元管理与规范框架为主要研究方向。他们的一项重要工作就是制定数据元标准：ISO/IEC 11179 Information Technology -- Metadata registries (MDR)。

ISO/IEC 11179 规范了描述数据所需数据元的种类和性质，并规定了如何利用数据元注册表（MDR）管理数据元。ISO 11179 标准共分为 6 个部分：

第 1 部分：框架。介绍并讨论了数据元、值域、数据元概念、概念域、分类表等基础理念，它们对于理解这一组标准起到关键作用。

第 2 部分：分类。提供了一个管理分类表的概念模型，并给出了分类表从两个层面的分类。

第 3 部分：注册元模型与基本属性。规定了数据元注册表的概念模型，以及当不需要一个完整的注册表时，描述数据元所用的一组基础属性。

第 4 部分：数据定义公式。提供开发无歧义数据定义的指南，严格规定了数据定义的格式。

第 5 部分：命名与标识原则。提供为管理中的条目设置标识的指南。

第 6 部分：注册。为数据项目的注册和标识设置提供了指导，同时还规定了注册项目的维护和管理工作。

(3) 美国环保局

美国环保局（EPA）负责研究和制定各类环境计划的国家标准，并且授权给州政府和美国原住民部落负责颁发许可证、监督和执行守法。在数据标准方面，EPA 提供稳定的格式化的数据元和数据取值集合。EPA 即采用得到广泛认可的现有标准，也自主制定一些标准。目前 EPA 认可的数据元标准包括生物学分类数据元、部落标识数据元、化学鉴定数据元、联系信息数据元、经纬度数据元、许可信息数据元等。EPA 还计划制定更多的数据标准。

EPA 数据标准中的数据元都按照 ISO 11179 的规定，注册到 EPA 数据注册服务（Data Registry Services）中。

EPA 数据注册服务还支持检索和下载数据元，从而支持环境数据的管理与使用。这些数据元描述了各机构数据系统中各字段的名称、定义、含义等信息。EPA 数据注册服务不包含任何环境数据，而是提供描述信息，使数据更有意义。

EPA 数据注册服务主要向用户提供：EPA 主要信息集数据字典、EPA 标准数据元的名称、定义、格式信息、EPA 代码表和相关的工具等。EPA 数据注册服务支持用户插入或维护数据字典、数据标准和代码表，鼓励用户为数据元定义补充概念和含义。

2、国内数据集相关资料

(1) 卫生领域

《中共中央国务院关于深化医药卫生体制改革的意见》和《国务院关于印发医药卫生体制改革近期重点实施方案（2009—2011 年）的通知》中明确提出：大力推进医药卫生信息化建设。数据作为医药卫生信息化建设的核心，其标准化工作也已取得了若干成果，如：电子病历系列标准、健康档案系列标准等。

“电子病历系列标准”：电子病历系列标准中涉及电子病历基本架构与数据标准、电子病历数据组与数据元标准和电子病历基础模板数据集标准。其中，《电子病历标准中涉及到电子病历基本架构与数据标准》包括电子病历的基本概念和体系架构、电子病历的基本内容和信息来源、电子病历数据标准等内容。《电子病历数据组与数据元标准》是我国电子病历数据标准的组成部分之一。本数据集旨在统一和规范电子病历的信息内涵，指导电子病历数据库及相关电子病历信息系统的开发设计，支持电子病历与相关卫生服务活动以及其它信息资源库相互间的数据交换与共享；同时为相关卫生服务活动的信息管理规范化与标准化提供依据，为构建整体的卫生信息模型和国家卫生数据字典提供基础信息资源。《电子病历基础模板数据集标准》包括《电子病历基本数据集编制规范》和 19 个电子病历基础模板数据集标准。

《健康档案基本架构与数据标准》：健康档案的建设是为了满足自我保健的需要、满足健康管理的需要、满足健康决策的需要。健康档案基础框架与数据标准的建设则是满足以上各类需求的基础。健康档案基本架构与数据标准包括《健康档案基本架构与数据标准》、《健康档案基本数据集编制规范》、《健康档案公用数据元标准》和 32 个基本数据集标准。

（2）人口计生领域

2007年，国家人口和计划生育委员会开展了人口宏观管理与决策信息系统（PADIS）项目的建设。该项目立足人口计生部门健全高效的组织体系和信息网络优势，依托国家电子政务网络平台，依靠制度化的信息交流与共享机制，与公安、统计、人口计生、卫生、财税、教育、劳动与社会保障、资源、环境、农业、建设等部门以及国际国内有关组织机构密切合作，系统收集、整合和利用现存的人口及相关经济社会宏观信息资源，建立信息共享、知识挖掘、决策支持、协调高效、安全稳定、保障有力的人口宏观管理与决策信息系统（PADIS）。

PADIS 系统建设十分重视标准化工作，建立了人口宏观管理与决策信息系统标准体系。该体系分为总体标准、管理标准、应用标准和应用支撑标准四方面 22 项标准。其中，应用标准强调以信息资源为核心，《人口和计划生育基础数据标准体系》便属于 PADIS 标准的应用类标准。

在《人口和计划生育基础数据标准体系》中规定了 PADIS 系统一期建设涉及的流动育龄妇女信息、计划生育家庭奖励扶助（救助）信息、人口快速调查与监测信息和人口计生事业信息的数据集划分，并以元数据的形式详细描述了每一个数据集的基本信息。

（3）国内其他标准情况

在进行国内调研的过程中，我们发现在对数据集进行说明时，都需要对数据集的数据元进行描述，因此，我们又调研了 GB/T 18391《信息技术 数据元的规范与标准化》系列标准、科学数据共享工程《数据元标准化的基本原则与方法》以及气象领域的《气象基础数据元目录》。

3.2 标准编制的原则

《环境数据集说明文档格式》的制定遵循了以下原则：

（1）本标准的编制与管理遵循国家环境保护总局 2006 年 8 月 31 日公布的《国家环境保护标准制修订工作管理办法》。

（2）按照规范的标准编制过程开展标准研制工作，包括项目开题及总体方案编制、资料调研、征求意见稿编制、送审稿编制、报批稿编制和质保期阶段。

（3）标准的编写严格按照 GB/T 1.1-2009《标准化工作导则 第 1 部分：标准的结构和编写》的要求。

4 标准主要技术内容

4.1 标准适用范围

本标准规定了环境数据集说明文档应包含的通用结构、内容要求、格式要求等，并规定了环境数据集说明文档模板。

本标准适用于环境数据集加工、使用、汇交过程中对环境数据集的说明。

4.2 标准结构框架

本标准共有 6 章和 2 个附录组成，主要内容如下：

第一章为适用范围，概述了本标准的适用范围。

第二章为规范性引用文件，表明了本标准内容所引用的文件中的条款。

第三章为术语和定义，列出了在本标准中出现的相关术语及其定义。

第四章为概述，环境数据集说明文档编写的总体要求。

第五章为环境数据集说明文档的通用结构和内容要求，提出了环境数据集说明文档的总体要求。并详细介绍了环境数据集说明文档应包含的基本章节，包括：概述、数据集标识、数据内容和结构、覆盖范围、数据质量、数据分发、限制信息、数据维护和元数据。

第六章为环境数据集说明文档格式要求，规定了环境数据集说明文档应遵循的格式要求。

规范性附录 A 为环境数据集说明文档模板，提供了一个环境数据集说明文档的模板。

资料性附录 B 以“2010 年 7 月 8 日重点城市空气质量日报数据集”为例，给出了环境数据集说明文档的参考性示例。

5 对实施本标准的建议

实施《环境数据集说明文档格式》标准，首先应确定要说明的对象，然后按照标准中的规定撰写说明文档。

(1) 明确说明对象

按照《环境数据集说明文档格式》标准的规定，文档说明的对象是环境数据集。环境数据集是由环境保护业务相关的若干数据实例组成、遵循统一的数据模型、具有相对独立的业务语义的数据集合。

环境数据集的粒度可大可小，应根据具体的应用需求来确定。例如，“2010 年 7 月 8 日重点城市空气质量日报数据”可以作为一个数据集；同样，根据业务应用需要，还可将 2010 年 7 月共计 30 天的重点城市空气质量日报数据组成一个更大粒度的数据集——“2010 年 7 月重点城市空气质量日报数据集”。

(2) 撰写说明文档

确定好说明对象之后，就可以遵循本标准撰写数据集的说明文档。撰写说明文档时，可按照本标准附录 A 中规定的模板，并参考附录 B 中给出的示例。对于文档各章节的具体内容要求，应遵循本标准第 5 章的规定；而行文格式的要求，应遵循本标准第 6 章的规定。

同时，本标准规定，在环境数据集说明文档中，应附上该数据集的元数据 XML 文件。有关元数据的要求，应遵循 HJ XXX-20XX《环境信息元数据标准》。

意见汇总处理表

2010年07月23日

序号	页码	章节	意见内容	提出人	处理意见	备注
1.	I	封皮	封皮标准名称“环境数据集说明文档格式标准”去掉“标准”二字； 封皮底部的“2010-□□-□□发布”、“2010-□□-□□实施”、“环境保护部 发布”格式调整。	刘定	采纳	标准名称调整为：“环境数据集说明文档格式”
2.	I	封皮	建议：将标准中文名下面增加英文名称	刘立媛	采纳	补充英文名称为：Documentation format of environmental dataset
3.	II	前言	“本标准主要起草单位：”增加“环境保护部信息中心”。	刘定	采纳	调整为：“本标准主要起草单位：环境保护部信息中心、北京市倍思电子数据库工程公司”
4.	1		标准名称改为“环境数据集说明文档格式”。	刘定	采纳	调整为：“环境数据集说明文档格式”
5.	1	1 适用范围	将“并规定了环境数据集说明文档模板”改为“并给出了环境数据集说明文档模板”。	刘定	采纳	调整为：“并给出了环境数据集说明文档模板”。
6.	1	1 适用范围	数据集说明文档格式不仅适用于数据集的加工与汇交，也适用于数据集的使用	总集成	采纳	调整为：“本标准适用于环境数据集加工、使用及汇交过程中对环境数据集的说明。”
7.	1	1 适用范围	“本标准规定了环境数据集说明文档应包含的通用结构、内容要求、格式要求等，并给出了环境数据集说明文档模板。”，如已阐述全面，“等”字可去掉。	环境标准研究所	采纳	调整为：“本标准规定了环境数据集说明文档应包含的通用结构、内容要求、格式要求，并给出了环境数据集说明文档模板。”
8.	1	2 规范性引用文件	将“HJ/T XXX-20XX 环境信息元数据规范”改为“HJ XXX-20XX 环境信息元数据规范”。	刘定	采纳	规范性引用文件调整为： HJ XXX-20XX 环境信息元数据规范 并同时修改标准正文中的相应内容。
9.	1	3.1	3.1 数据集 dataset 可识别的数据集合。[GB/T19710-2005] 引用的标准[GB/T19710-2005]可在编制说明中给出，此处删除	环境标准研究所	部分采纳	在《标准化工作导则 第1部分：标准的结构和编写》（GB/T1.1-2009）6.3.2 术语和定义一节中要求“如果确有必要重复某术语已经标准化的定义，则应标明该定义出自的标准。”按照该节给出的示例1所示，同时给出所引用术语在标准中定义

						的具体章节。 修改为： 3.1 数据集 dataset 可识别的数据集合。 注：通过诸如空间范围或要素类型的限制，数据集在物理上可以是更大数据集较小的部分。从理论上讲，数据集可以小到更大数据集内的单个要素或要素属性。一张硬拷贝地图或图表均可以被认为是一个数据集。 [GB/T19710-2005，定义 4.2]
10.	1	3.2 环境数据集	建议将环境数据集的定义应该有环境的含义在里面，比如“环境保护业务相关的……”。	孙强	采纳	将环境数据集的定义调整为：“由环境保护业务相关的若干数据实例组成、遵循统一的数据模型、具有相对独立的业务语义的数据集合。”
11.	1	4 概述	将“该文档的通用结构和内容要求见第 5 章”去掉第一个“该”字	刘定	采纳	修改为：“文档的通用结构和内容要求见第 5 章”。
12.	1	5.1 总体要求	将“环境数据集说明文档应包含下列基本章节：”下面的 a) 到 e) 行句末的标点“。”改为“；”； 将“环境数据集说明文档还可包括以下栏目：”下面的 a) 到 c) 行句末的标点“。”改为“；”。	刘定	采纳	修改为： 环境数据集说明文档应包含下列基本章节： a) 概述——见 5.2 节； b) 数据集标识——见 5.3 节； c) 数据内容和结构——见 5.4 节； d) 覆盖范围——见 5.5 节； e) 元数据——见 5.10 节； 环境数据集说明文档还可包括以下栏目： a) 数据质量——见 5.6 节； b) 数据分发——见 5.7 节； c) 限制信息——见 5.8 节； d) 数据维护——见 5.9 节。
13.	1	5.1 总体要求	建议将“元数据”按照顺序安排在 5.6 节	环境标准研究所	未采纳	本节分别介绍了环境数据集说明文档中，哪些章节为必选的，哪些章节是可选的。元数据是必选章节，但按照文档内容的逻辑顺序，建议将元数据放在所有章节之后。
14.	2	5.2 概述	建议在“环境数据集说明文档的创建信息”中包含数据集的“数据来源”	总集成	部分采纳	“环境数据集说明文档的创建信息”中重点说明“环境数据集说明文档”的相关信息；在 5.3“数据集标识”中增加“数据集提供

						方”： d)数据集提供方：包括数据集提供单位、数据集提供方电话、数据集提供方电子邮件地址； 关于数据集来源的其他信息，可在 5.6“数据质量”中的“数据志说明”中加以阐述。
15.	2	5.3 数据集标识	将“若需要可包括以下可选项：”下面的“a)目的：生成数据集的目的总结；”去掉句末的“总结”	刘定	采纳	修改为：“a)目的：生成数据集的目的”。
16.	2	5.4.1 数据模型	建议：UML 可以规范的描述数据模型，但如果是数据集只是一个数据库文件、一张报表、一个专题图，如果用 UML 描述数据内容？建议增加数据格式、基本内容描述的说明项。	孙强	采纳	本标准中已规定了“数据分发格式”（见 5.7 的 a)) 以对数据格式进行描述；对于数据基本内容的描述，可以采用“数据集摘要”（见 5.3 的 b)) 实现对数据内容的概要描述。
17.	3	5.4.1 数据模型	将本节“d)角色”下面的“图 5 说明在 UML 模型中如何表示角色名称和基数。”中的“图 5”改为“图 4”。	刘定	采纳	改为“图 4 说明在 UML 模型中如何表示角色名称和基数。”。
18.	3	5.4.1 数据模型	c)，解释超类和泛化类	环境标准研究所	采纳	调整为：泛化表示一般化描述和具体化描述之间的关系，即超类与可以替代它的子类之间的关系。
19.	3	5.4.2 数据元	关于数据元部分要与三个数据元标准保持一致；	总集成	采纳	
20.	3	5.4.2 数据元	将本节第三行“则应直接引用”改为“可直接引用”。	刘定	采纳	改为：“若所包含的数据元在《污染源监督性监测数据元技术规范》、《污染源在线监控数据元技术规范》、《环境统计数据元技术规范》中已有定义，可直接引用”
21.	3	5.4.2.2	“b)要阐述其概念是什么，而不是阐述其概念不是什么。说明：编写定义时，仅阐述其概念不是什么并不能对概念作出唯一的定义。”建议语言书面化，避免口语化	环境标准研究所	采纳	调整为：b)要阐述其概念的准确含义，避免歧义性的定义。
22.	5	5.4.2.4 值域	将本节最后一行“代码表可以被多个数据元使用，即可被再利用。”改为“代码表可以被多个数据元使用，可被再利用。”。	刘定	采纳	改为：“代码表可以被多个数据元使用，可被再利用。”。
23.	5	5.4.2.3 数据格式	本节倒数第二行“例 3: Ture/False 表示该数值为布尔型。”中修改“Ture”为“True”。	刘立媛	采纳	检查全文拼写错误并修改。
24.	5	5.4.2.5	“代码表模版如下”调整为“代码表模版如表 2 所示”	环境标准	采纳	调整为“代码表模版如表 2 所示”

		数据格式		研究所		
25.	5	5.4.2.	将表格内容的字号调为小五。	环境标准研究所	采纳	
26.	6	5.5 覆盖范围	建议“空间覆盖范围”可以进一步详细到经纬度； 建议时间范围不仅包括“数据内容跨越的时间段,包含起始时间和结束时间”还包括“时间频度”	总集成	采纳	1、在“空间覆盖范围”中增加可选项：经度、纬度。修改为： a) 空间覆盖范围：包括空间范围、经度和纬度。其中，空间范围用以说明数据内容覆盖的空间区域范围，可多选，如流域名称、全国或省市名称，其中省市名称按 GB/T 2260-2002 执行，按照“表 1 省、自治区、直辖市、特别行政区代码表”中的“名称”取值；经度、纬度为可选项。 2、增加时间频度，以说明该数据集采集的频率。
27.	8	附录 A	对附录 A 标题“环境数据集说明文档模板”提出以下建议：此处标题中不应出现环境字样，就是数据集说明文档。	孙强	采纳	附录 A 标题仍保留“环境数据集说明文档模板”，调整文档模板封面标题为“XXXX 数据集说明文档”
28.	11	附录 A	建议：增加附录 B（资料性附录）XXX 环境数据集说明文档。	刘定 孙强	采纳	增加附录 B“环境数据集说明文档示例”
29.		文档页眉	调整文档页眉格式，将“HJ□□-20□□”调整为：“HJ□□□-20□□”	刘定	采纳	将页眉调整为：HJ□□□-20□□