

附件 3

《环境与健康数据字典（第一版）
（征求意见稿）》编制说明

《环境与健康数据字典》编制组

二〇一八年二月

项目名称	环境与健康数据字典（第一版）
项目编制单位	环境保护部环境与经济政策研究中心、中国环境科学研究院、中国环境科学学会
编制组主要成员	王建生、韦正峥、黄炳昭、张金良、赵秀阁、邢晶晶、朱忠军、陈昱
环境保护部科技标准司环境健康管理处项目负责人	宛悦、陈子易

目 录

1	项目背景.....	1
2	必要性和可行性.....	1
3	国内外研究现状.....	2
4	编制原则.....	3
5	主要说明.....	7

1 项目背景

1.1 任务由来

发布《环境与健康数据字典》为 2018 年度环境保护部环境与健康工作任务，编制任务由环境保护部环境与经济政策研究中心承担。

1.2 工作过程

2017 年 1 月至 3 月：环境保护部科技标准司向环境保护部环境与经济政策研究中心下达了《环境与健康数据字典》编制任务。环境保护部环境与经济政策研究中心成立了由中国环境科学研究院、中国环境科学学会有关专家组成的编制组，并组织召开了《环境与健康数据字典》开题论证会，明确了数据字典的编制原则和基本框架。

2017 年 4 月：编制组在前期研究的基础上，形成了《环境与健康数据字典（初稿）》。

2017 年 4 月至 10 月，编制组针对《环境与健康数据字典（初稿）》邀请不同专业专家进行了 30 余次讨论会，对数据字典收录的数据元进行了逐条讨论修改。

2017 年 11 月至 12 月，编制组按照《环境信息元数据规范》（HJ 720 - 2017）中提出的数据元的元数据和术语的元数据进行了表达形式的完善，并形成了《环境与健康数据字典（第一版）》（征求意见稿草案）及其编制说明。

2018 年 1 月：标准编制组邀请相关专家对标准征求意见稿草案进行了论证。编制组在充分吸取专家意见后形成了《环境与健康数据字典（第一版）》（征求意见稿）及其编制说明。

2 必要性和可行性

2.1 必要性

随着我国环境污染导致人群健康危害问题的日益突出，开展环境健康风险管理已经成为环保部门的一项重要工作。环境健康风险管理必须以大量基础数据作为支撑。然而，由于相关数据标准建设滞后，特别是缺乏统一的数据字典和数据集标准规范，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，极大地限制了后续的数据共享使用，甚至出现了虽然有大量数据，却无法共享使用的尴尬局面。《国家环境保护“十三五”环境与健康工作规划》提出，我国要“建立环境与健康信息共享机制，规范信息发布方式”。

数据字典是按照一定的顺序、一定的规则和内容说明方式对数据的最小组成单位进行描述，并集合了关于该数据的所有信息，属于元数据范畴，由实体和数据元组成。数据字典是对数据的规范化描述，能够准确地说明数据的各项特征。通过数据字典，能够使用户对信息的定义、取值、格式、质量、处理方法和获取方法等各方面有更深入、细致的了解，从而实现数据语义的统一。环境与健康数据字典是相关利益人一致认可的对数据的一致性表达，包括定义、取值、格式、处理方法和获取方法等，目的是从源头保证不同来源的环境与健康数据能够有准确、

一致的理解和表达，从而增加环境与健康领域信息的可比性和可利用性，为今后建立统一高效、资源整合、互联互通和信息共享的环境与健康信息化网络奠定基础。

2.2 可行性

2012—2017 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，编制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，并开始指导实际业务工作中的信息采集。同时，结合重点地区环境与健康调查、暴露参数调查等数据管理需要，研究成果在实际工作中得到成功应用。上述业务实践，为编制《环境与健康数据字典》奠定了基础。同时，环境保护部于 2017 年修订了《环境信息元数据规范》(HJ 720—2017)，增加了数据元和术语的元数据规范，为本字典的表达提供了依据。

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国家总体行为之中。

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)，为国家卫生数据字典 (NHDD, National Health Data Dictionary, Version 16)、国家住房数据字典 (NHHDD, National Housing and Homelessness Data Dictionary, Version 1)、国家社区服务数据字典 (NCSDD, National Community Services Data Dictionary, Version 8) 提供注册服务。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 2 600 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。美国环境保护局发布了包括化学物质登记系统、清洁流域需求调查、空气质量系统等在内的 18 个数据字典，每个数据字典由表构成，表由列构成。其中，数据字典就类似于相关内容或主题的一组数据集规范的集合，表类似于数据集规范或基本数据集，列相当于变量，类似于数据元素。

3.2 国内研究情况

我国数据元标准化的研究工作始于 20 世纪 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”。

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，并已开放使用。《国家卫生数据字典》是卫生领域的元数据资源库，由数据元及其它元数据组成，主要用途是从国家层面指导卫生行业不同领域、不同机构、不同信息系统卫生信息的收集和利用，为居民电子健康档案和电子病历的跨机构共享提供基本条件。目前，《国家卫生数据字典》包含了个人基本信息和来自儿童保健、妇女保健、疾病预防控制、疾病管理及医疗服务等业务领域的相关信息项目，共1465个数据元，178个值域。基本数据集是数据字典的应用之一，是国家卫生数据字典的附属结构。医药卫生领域目前已发布《城乡居民健康档案基本数据集》（WS 365-2011）、《疾病管理基本数据集》（WS 372-2014）、《疾病控制基本数据集》（WS 375-2014）、《儿童保健基本数据集》（WS 376-2014）、《妇女保健基本数据集》（WS 377-2014）、《电子病历基本数据集》（WS 445-2014）等180余个基本数据集。将数据字典的功能进一步深化应用。

2002年，国土资源部信息中心发布《国土资源信息核心元数据元素指南》，起到数据字典的功能。该指南对核心元数据的内容、组织方式的简要说明外，主要内容是结合实例对每一项元数据元素的定义、内容描述、约束条件、最多出现次数、数据类型、值域等进行描述。特别是值域部分，比较详细地讨论了元素的典型取值，并给出相应的实例，有助于用户在进行数据采集时作出正确的选择。

2010年，住房和城乡建设部发布《城市地理空间信息共享与服务元数据标准》（CJJ/T 144-2010），规定了元数据的内容、获取、管理与发布。附录A为“元数据字典”，对元数据的实体信息、标识信息、内容信息等166个数据元在中英文名称、定义及表示格式上均做了详细的描述。

2017年，中国气象局发布《气象数据集核心元数据》（GB T 33674-2017），规定了完整描述一个气象数据集时所需要的数据项集合、各数据项语义定义和著录规则等，它提供了有关气象数据集的标识、内容、分发、数据质量、数据表现、参照和限制等信息。附录A为“气象数据集核心元数据字典”，对时间标识、地理覆盖范围、垂向覆盖范围、时间覆盖范围等信息的88个数据元在中英文名称、定义及表示格式上均做了详细的描述。

2017年，环境保护部修订并发布了《环境信息元数据规范》（HJ 720-2017），调整了元数据管理的思路，对对象类、特性、分类方案、值域、数据元概念、数据元、数据集规范、术语、指标、数据集、质量声明共11个管理项进行规范。其中数据元和术语的元数据规范为编制数据字典提供了依据。

4 编制原则

4.1 以《环境信息元数据规范》为依据进行表达

本数据字典是按照《环境信息元数据规范》（HJ 720—2017）中确定的数据元和术语的元数据项目，对环境与健康调查、监测及风险评估领域常用数据元和术语进行标准化。其中数据元的元数据包括6类属性：标识类、定义类、表示类、关系类、管理类、附加类。各类属性详细描述见表1。

由于标识类和管理类属性只用于数据标准的注册管理，而不涉及数据定义和表达的统一，因此在本数据字典中予以省略，以方便用户阅读。最终本数据字典

中用于描述数据元的元数据属性共保留了 5 个属性中的共 16 个字属性，具体以数据元“土壤样品-镉浓度，【计量|nnnn.nn|mg/kg】”为例，见下表 2。

术语包含 4 类 16 项基本属性。其中由于标识类和管理类属性只用于数据标准的注册管理，而不涉及数据定义和表达的统一，因此在本数据字典中予以省略，以方便用户阅读。具体以术语“暴露量”为例，见下表 3。

表 1 数据元的属性

属性种类	属性名称	约束
标识类	中文名称	M
	英文名称	M
	变量名	M
	标识符	M
	版本	M
	元数据项类型	M
	提交机构	M
	同义名称	O
	相关环境	C
定义类	定义	M
	相关数据元概念	M
	相关值域	M
表示类	表示类别	M
	数据类型	M
	表示格式	M
	最小长度	M
	最大长度	M
	允许值	M
	计量单位	C
关系类	使用指南	O
	标准引用	O
	评价标准	O
	相关数据集规范	O
	相关数据集	O
	相关指标	O
	相关术语	O
管理类	注册机构	M
	注册状态	M
	注册日期	M
	备注	O
附加类	采样方法	O
	检测方法	O

表 2 数据元的规范化描述示例

标识类	
中文名称:	土壤样品的镉浓度
英文名称:	Soil Samples-Cd Concentration
元数据项类型:	数据元
定义类	
定义:	单位质量土壤中镉的含量。
关系类	
使用指南:	镉是人体非必需元素，在自然界中常以化合物状态存在，一般含量很低，正常环境状态下，不会影响人体健康。当环境受到镉污染后，镉可在生物体内富集，通过食物链进入人体引起慢性中毒，损伤肝、肾、骨骼的功能，典型的如镉大米。采集含镉土壤样品可选择聚乙烯或玻璃瓶储存。
标准引用:	《展览会用地土壤环境质量评价标准（暂行）》（HJ 350—2007）； 《温室蔬菜产地环境质量评价标准》（HJ 333-2006）； 《土壤环境监测技术规范》（HJ/T166-2004）； 《土壤质量 铅、镉的测定 石墨炉原子吸收分光光度法》（GB/T17141-1997）； 《土壤和沉积物 金属元素总量的消解 微波消解法》（HJ 832-2017）
评价标准:	《土壤环境质量标准》（GB15618—1995）：一级标准，土壤中镉浓度限值为 0.2mg/kg；二级标准，当 pH<6.5 时，土壤中镉浓度限值为 0.3mg/kg；当 pH 介于 6.5~7.5 之间时，土壤中镉浓度限值为 0.3mg/kg；当 pH>7.5 时，土壤中镉浓度限值为 0.6mg/kg；三级标准，pH>6.5 时，土壤中镉浓度限值为 1mg/kg。 《展览会用地土壤环境质量评价标准（暂行）》（HJ 350—2007）：A 级标准镉浓度限值为 1mg/kg，B 级标准镉浓度限值为 22mg/kg。 《食用农产品产地环境质量评价标准》（HJ 332-2006）：当土壤用于水作、旱作、果树等，总镉浓度限值为 0.3mg/kg（pH≤7.5）及 0.6 mg/kg（pH >7.5）；当土壤用于种植蔬菜时，总镉浓度限值为 0.3mg/kg（pH≤7.5）及 0.4mg/kg（pH >7.5）。 《温室蔬菜产地环境质量评价标准》（HJ 333-2006）：总镉浓度限值为 0.3mg/kg（pH≤7.5）及 0.4 mg/kg（pH >7.5）。
表示类	
表示类别:	计量
数据类型:	数值型
表示格式:	nnnn.nn
最小长度:	4
最大长度:	7
允许值:	大于零的实数
计量单位:	mg/kg
附加类	
采样方法:	《土壤环境监测技术规范》（HJ/T166-2004）
检测方法:	《土壤质量 铅、镉的测定 石墨炉原子吸收分光光度法》（GB/T17141-1997）； 《土壤和沉积物 金属元素总量的消解 微波消解法》（HJ 832-2017）

表 3 术语的规范化描述示例

标识类	
中文名称:	暴露量
英文名称:	Exposure Dose
元数据项类型:	术语
定义类	
定义:	指人体经呼吸道、消化道和皮肤等途径接触环境污染物的量。
关系类	
相关术语:	暴露、暴露量、暴露途径
标准引用:	《暴露参数调查技术规范》(HJ 877—2017)
相关元数据项:	调查对象-血铬浓度,【计量 nn.nnn μg/L】 调查对象-血镉浓度,【计量 nn.nn μg/L】 调查对象-尿镉浓度,【计量 nn.nn μg/g Cre.】 调查对象-尿砷浓度,【计量 nn.nn μg/g Cre.】 调查对象-发砷浓度,【计量 nn.nn μg/g】 调查对象-血铅浓度,【计量 nnnn.nn μg/L】 调查对象-血锌原卟啉(ZPP)浓度,【计量 nn.nn μmol/L】 调查对象-尿汞浓度,【计量 nn.nn μg/g Cre.】 调查对象-发甲基汞浓度,【计量 nn.nn μg/g】 调查对象-尿液苯并[a]芘浓度,【计量 nn.nn pg/g Cre.】

4.2 满足环境与健康业务应用为原则

本次发布的《环境与健康数据字典(第一版)》紧紧围绕着当前环境与健康工作的实际需求,在内容安排上,采取了成熟一批、纳入一批、发布一批的原则,包括了环境与健康领域常用数据元的名称、定义、表达格式及使用指南,以期尽快的启动数据标准化工作。随着今后环境与健康工作的不断扩充发展,本数据字典的内容也将会根据实际业务需求进行扩充和更新。在确定纳入本数据字典中数据元和术语时,主要采取了以下原则:

1. 目前正在开展的环境与健康专项调查、环境与健康哨点监测、暴露参数调查、环境与健康素养调查等重点项目中涉及关键数据元优先纳入;
2. 目前已经有成熟标准的数据元优先纳入;
3. 目前环境保护部发布的标准中设计的术语优先纳入。

4.3 与现有数据标准兼容原则

目前卫生系统已经发布了几十项的基本数据集标准,其中涵盖了大量的数据元,以及值域代码。在环保系统,也发布了一些环境信息代码类标准。因此,在本数据字典编制过程中,编制组参考了《废水排放去向代码》(HJ 523-2009)、《中国地表水环境水体代码编码规则》(HJ 932-2017)、《城乡居民健康档案基本数据集》(WS365-2011)、《卫生信息数据元值域代码》(WS 364-2011)等资料,凡是在卫生

信息标准或环境信息标准中已经涵盖的数据元或者值域代码，则按照现有标准中的规范进行引用。在使用指南和附加属性中尽可能补充。

4.4 尽量做到科学权威的原则

数据字典编制过程中，编制组参考了国家统计局国民经济核算指标解释、第六次全国人口普查、中国气象数据中心-专业词汇解释、全国污染源普查工业源普查详表指标解释与填报说明、《环境监测技术规范》、各类检测方法相关技术文件、《中国人群暴露参数手册》等资料，在资料的使用中，注重其权威性和时效性，确保资料的可信度。

并且，在质量控制方面，本数据字典中收录的所有数据元都经过中国环境科学学会环境与健康数据标准与信息共享专家委员会组织的论证和审核。

5 主要说明

1. 本数据字典内容包括了数据元和术语两部分内容，共 249 条。其中数据元的排列按照用户的使用习惯进行排列，分为 9 个部分，包括：

(1) 社会经济状况相关数据元 (25 条)：含人口、经济、气象条件信息等；

(2) 污染源相关数据元 (47 条)：含污染源登记信息、废气、废水、废渣现场采样信息等；

(3) 环境质量相关数据元 (47 条)：含环境水、环境空气、土壤采样信息等；

(4) 暴露测量相关数据元 (20 条)：含家庭饮用水、室内空气、积尘、农作物、禽畜产品、水产品采样信息等；

(5) 暴露参数相关数据元 (32 条)：含粮食蔬菜摄入量、饮水摄入量、吸烟饮酒情况等；

(6) 个体基本信息相关数据元 (6 条)：含性别、职业、本地居住年限等；

(7) 污染物人体内负荷相关数据元 (11 条)：含血、尿、发中的汞、镉、铅、铬、砷等重金属的含量；

(8) 健康效应评估相关数据元 (11 条)：含肾功能、肿瘤、死亡情况等；

(9) 术语 (50 条)：将目前环保部发布的 5 个环境与健康相关标准涉及术语全部著录，包括《儿童土壤摄入量调查技术规范 示踪元素法》(HJ 876—2017)、《环境污染物人群暴露评估技术指南》(HJ 875—2017)、《环境与健康现场调查技术规范 横断面调查》(HJ 839—2017)、《人体健康水质基准制定 技术指南》(HJ837-2017)、《暴露参数调查技术规范》(HJ 877—2017)。

2. 为了方便用户根据首字母查阅，我们在数据字典最后附“A-Z 索引列表”。

3. 本数据字典的表达按照《环境信息元数据规范》(HJ-720 2017)中的元数据项目进行表达，其中由于标识类和管理类属性只用于数据标准的注册管理，而不涉及数据定义和表达的统一，因此在本数据字典中予以省略，以方便用户阅读。

4. 由于数据元由对象类、特性和表示三部分构成，因此本数据字典在数据元表达时分别涵盖了对对象类、特性和表示的主要内容。例如对于“土壤样品-铅浓度，【计量|nnnn.nn|mg/kg】”这个数据元，其中“土壤样品”为对象类；“铅浓度”为描述土壤样品的特性；“【计量|nnnn.nn|mg/kg】”为这个数据元的表示方式，其中“计量”表示这个数据元为计量型资料，“nnnn.nn”表示这个数据元的格式为四位整数加两位小数，“mg/kg”表示这个数据元的度量衡单位必须为毫克每千克。

5. 本数据字典中收录的数据元、术语，以及构成数据元的对象类、特性、数据元概念、表示均已经在“环境与健康数据字典及元数据注册系统”中进行了注册管理，用户可以通过在线访问 <http://www.eh-data.cn:8093/metadata/> 来检索查询。

6. 本字典引用的标准及规范均注明日期，若出现更新以最新版本为准，本字典也将持续更新。随着今后环境与健康工作的不断扩充发展，本数据字典的内容也将会根据实际业务需求进行扩充和更新。