

《生态环境大数据数据字典技术规范》

(征求意见稿)

编制说明

一、任务来源

科技部为贯彻落实《关于加快推进生态文明建设的意见》，按照《关于深化中央财政科技计划（专项、基金等）管理改革的方案》（国发〔2014〕64号）要求，制定了国家重点研发计划“场地土壤污染成因与治理技术”重点专项实施方案，设立了《基于人工智能与区块链技术的生态环境新型治理体系研究与示范应用》研究项目。北京市生态环境监测中心、北京环丁环保大数据研究院、清华大学等单位承担了这一项目。

自20世纪80年代中期以来，我国的生态环境监测信息化工作经历了30余年的发展，基本建成了要素上覆盖环境质量、自然生态、重点污染源等全部环境领域，空间范围上覆盖全部城市和区县的生态环境监测网络。环境监测信息化的发展为环境领域积累了海量数据，推动生态环境监测进入大数据时代。但是，在长期的环境监测实践中，因时间不同、负责部门不同、开发团队不同、采用的技术不同，环境监测部门主导开发或运行的多个环境监测系统相互之间数据不统一，对同一环境对象、业务过程的描述各不相同，各业务系统之间很难进行数据共享和业务协同。

为了进一步提升我国生态环境数据的管理水平和利用效率，确保数据的准确性、一致性和可共享性，打破“数据孤岛”和“应用烟囱”，

国家及社会各界对生态环境大数据的管理提出了更高要求。在此背景下，中华环保联合会积极响应国家号召，联合北京环丁环保大数据研究院、北京市生态环境监测中心、北京微芯区块链与边缘计算研究院、联通数字科技有限公司等多家科研机构及企业，共同承担了《生态环境大数据数据字典构建技术规范》团体标准的编制任务。

该任务不仅旨在规范生态环境大数据的数据字典构建过程，提高数据质量，更希望通过统一的技术标准，推动全国范围内生态环境数据的整合与共享，为政府决策、环境监管、生态保护提供有力支撑。这一标准的制定，是国家生态文明建设和环境保护工作的重要一环，也是推动我国生态环境保护事业迈向新台阶的关键举措。

二、目的、意义、必要性

1. 目的

制定《生态环境大数据数据字典构建技术规范》的核心目的，在于明确生态环境大数据数据字典的内容组成和属性填写要求，为生态环境大数据平台的设计、开发、管理和维护提供统一的技术规范。这不仅能够确保数据在收集、存储、处理、分析等各个环节中的一致性和准确性，还能够促进不同系统、不同部门之间的数据共享与交换，提高数据的整体利用价值。

2. 意义

提升数据质量：通过统一的数据字典构建标准，减少数据冗余和错误，提高数据的准确性和可靠性。

促进数据共享：打破信息孤岛，实现跨部门、跨地区的数据互通

共享，为生态环境保护工作提供更加全面、详实的数据支持。

支持科学决策：高质量的数据是科学决策的基础。通过标准化的数据字典构建，为政府决策提供更加精准、客观的数据依据。

3. 必要性

在当今社会，随着科技的飞速发展，我们越来越依赖于数据来洞察世界、指导决策。在生态环境领域，大数据的应用更是不可或缺。然而，面对海量、复杂且多变的生态环境数据，如何有效地管理、整合和利用这些数据，成为了一个亟待解决的问题。在此背景下，制定“生态环境大数据数据字典构建技术规范”显得尤为重要，其必要性主要体现在以下几个方面：

(1) 保障数据的一致性与准确性

生态环境数据来源于多个领域、多个渠道，数据类型多样，格式不一，这就导致了数据在整合过程中可能出现的混乱和误差。数据字典作为数据的“说明书”，详细记录了数据的每一个细节，包括数据结构、数据类型、数据含义、数据来源等。通过制定技术指南，可以规范数据字典的构建流程，确保不同数据源之间的数据在整合时能够保持一致性，减少因数据格式、定义不清等问题导致的错误，从而提高数据的准确性。

(2) 促进数据的共享与协作

在生态环境保护工作中，数据共享与协作是提高工作效率、实现协同治理的关键。然而，由于数据标准的缺失，不同单位、不同部门之间在数据交换时常常面临困难。制定技术指南，可以明确数据字典

的构建标准和要求，使得各单位在构建数据字典时遵循相同的规范，从而便于数据的共享与协作。这不仅能够减少重复劳动，降低工作成本，还能够促进跨学科、跨领域的合作与交流，推动生态环境保护工作的深入开展。

(3) 提升数据的利用价值

生态环境大数据蕴含着丰富的信息和价值，但要想充分挖掘这些价值，就必须先对数据进行有效的管理和分析。数据字典作为数据管理与分析的基础工具，其质量直接影响到数据的利用效果。制定技术指南，可以确保数据字典的构建质量，提高数据的可读性和可理解性，使得数据在后续的分析、挖掘和应用中能够更加高效、准确。这不仅能够提升数据的利用价值，还能够为生态环境保护提供更加科学、有力的决策支持。

综上所述，制定“生态环境大数据数据字典构建技术规范”是保障数据一致性与准确性、促进数据共享与协作、提升数据利用价值的必然要求。该标准的出台，将为生态环境大数据的管理和利用提供有力的技术支撑和制度保障，推动生态环境保护事业的持续健康发展。

三、主要工作过程及标准编制技术路线

在《生态环境大数据数据字典构建技术规范》的编制过程中，我们遵循了严谨、科学、民主的原则，确保了标准的适用性和可操作性。

1. 主要工作过程

国内外现状调研：我们广泛收集了国内外关于生态环境大数据管理的相关资料，包括技术标准、应用案例、政策法规等，深入分析

了当前生态环境大数据管理的现状、存在的问题及发展趋势。

需求分析：基于调研结果，我们明确了生态环境大数据数据字典构建的实际需求，包括数据一致性、准确性、共享性的要求，以及生态环境保护工作的实际需求。

组建起草小组：我们邀请了来自北京环丁环保大数据研究院、北京市生态环境监测中心、北京微芯区块链与边缘计算研究院、联通数字科技有限公司等单位的专家和技术人员，共同组建了标准起草小组，起草组名单见表 1。

明确分工：小组成员根据各自的专业领域和擅长方向进行了明确分工，确保了标准编制工作的顺利进行。

表 1 起草组人员名单

编号	姓名	单位	标准编制分工
1	吕广丰	北京环丁环保大数据研究院	主持人
2	王小菊	北京市生态环境监测中心	框架、章节起草
3	杨懂艳	北京市生态环境监测中心	章节起草
4	马俊文	北京市生态环境监测中心	章节起草
5	邱昀	北京市生态环境监测中心	章节起草
6	黄瑾	北京微芯区块链与边缘计算研究院	章节起草
7	庄缘	北京联通数字科技有限公司	全文统筹
8	王露露	北京环丁环保大数据研究院	校对

技术框架设计：在充分调研和需求分析的基础上，我们设计了标

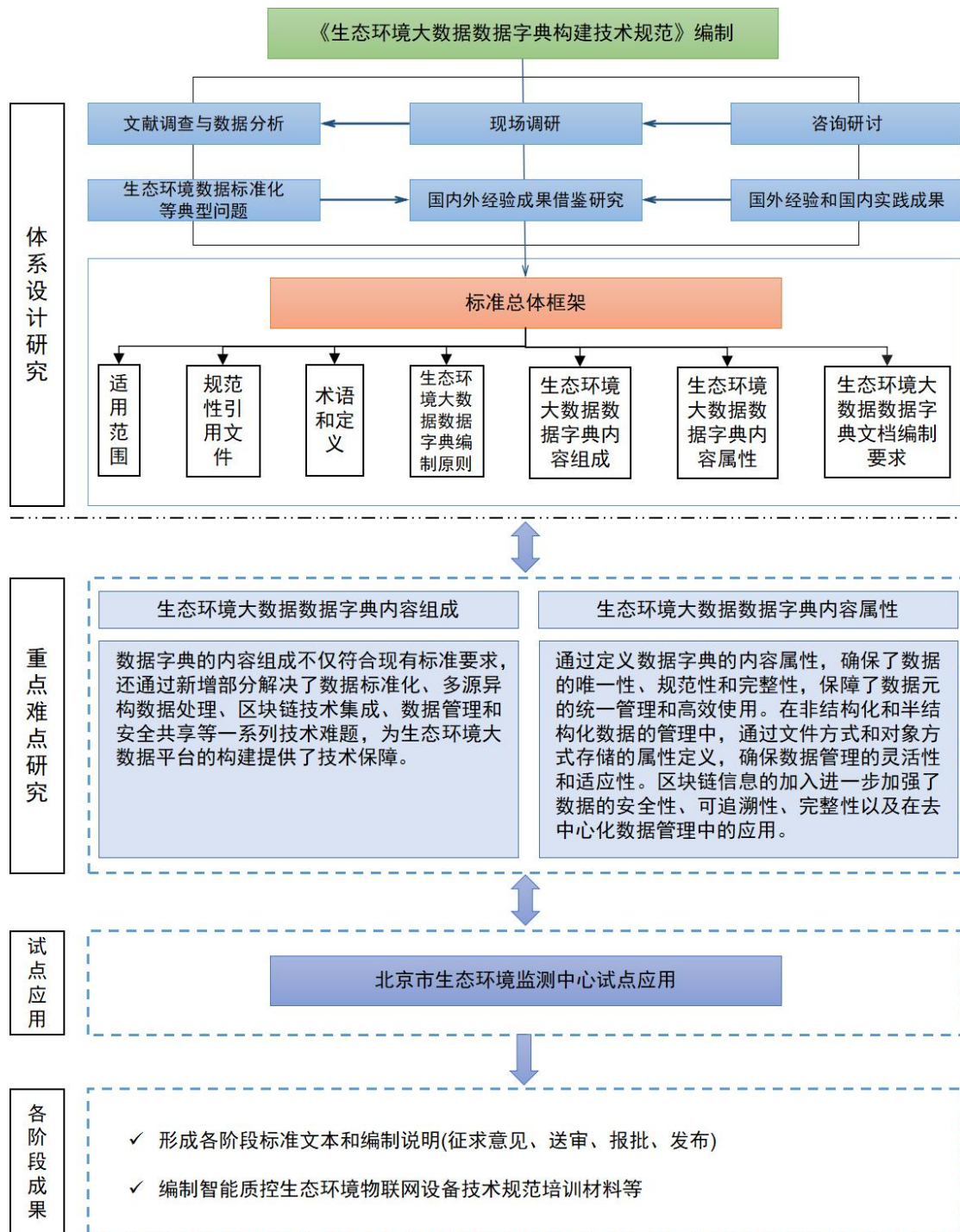
准的技术框架，明确了标准的主要内容、结构布局和编写要求。

编写初稿：起草小组根据技术框架，结合国内外相关标准和最佳实践，编写了标准的初稿。初稿涵盖了生态环境大数据数据字典的构建原则、内容组成、属性填写要求等方面。

召开立项会：2024年6月5日，标准立项会召开。会议评估了拟立项标准的必要性和可行性，专家们经过讨论一致认为该项目符合立项条件。

修改完善：根据立项会专家评审意见，我们对初稿进行了多次修改和完善，确保标准的科学性和合理性。

2. 标准编制技术路线



四、同类标准国内外状况

1. 国内外同类标准状况

在国内，随着生态文明建设的深入推进，生态环境大数据的管理

和应用也日益受到重视。近年来，国家相关部门已经出台了一系列政策文件和技术规范，旨在推动生态环境大数据的发展和应用。环境保护部也于 2014 年颁布《环境信息数据字典规范》（HJ 723-2014）行业标准。然而，在行标颁布时大数据技术商务推广开来，区块链技术尚未在业务系统中得到应用，行标已经不能适应新时期环境信息化的发展需求。

许多行业主管部门也颁布了本行业的数据字典规范，如国家林业局颁布的《林业信息数据库数据字典规范》（LY/T 2672-2016）行业标准，该标准规定了林业信息数据库及其内部各类数据的数据字典编写内容、格式、方法及其数据字典的注册要求和文件格式要求等内容。支撑了林业信息库的数据集成、管理、维护、分发服务和共享等业务。交通部颁布的《交通运输信息系统数据字典编制规范》（JT / T 1020-2016）行业标准，标准规定了交通行业数据字典内容、域含义和填写要求以及数据字典文档编制要求等内容，有效支撑了交通运输行业信息系统之间的互联互通和数据共享。

在国际上，随着大数据技术的快速发展，各国政府和行业协会已经开始重视大数据标准的建设工作。特别是在生态环境保护领域，一些发达国家如美国、欧盟等已经制定了较为完善的数据管理和共享标准体系。这些标准不仅关注数据的采集、处理和分析技术，还涉及数据的隐私权保护、数据安全等多方面内容。例如，欧盟的《通用数据保护条例》（GDPR）就对数据保护提出了严格要求，为生态环境保护大数据的管理提供了法律框架。

2. 同类项目现状

在生态环境大数据管理领域，国内外已经开展了一系列相关项目和研究工作。这些项目大多聚焦于数据的采集、存储、处理和分析等关键技术环节，旨在提升生态环境数据的管理水平和利用效率。例如，一些地方政府通过建设生态环境大数据平台，实现了对区域内环境质量的实时监测和预警；一些科研机构则致力于开发高效的数据处理算法和分析模型，为生态环境保护提供科学依据。

然而，在数据字典构建方面，这些项目大多依据各自的需求和技术路线进行开发，缺乏统一的标准和规范。这导致了不同项目之间数据难以共享和交换的问题，限制了生态环境大数据的整体应用效果。因此，制定统一的生态环境大数据数据字典构建技术规范显得尤为重要。

五、标准核心技术内容及说明

1. 标准结构框架

本标准主要包括适用范围、规范性引用文件、术语和定义、生态环境大数据数据字典编制原则、生态环境大数据数据字典内容组成、生态环境大数据数据字典内容属性及生态环境大数据数据字典文档编制要求等方面内容。

- (1) 适用范围：概述了本标准的编制内容和适用范围。
- (2) 规范性引用文件：介绍了本标准中引用的相关标准文件。
- (3) 术语和定义：规定了标准中的相关术语。
- (4) 生态环境大数据数据字典编制原则：规定了生态环境大数

据数据字典的编制原则。

(5) 生态环境大数据数据字典内容组成：规定了生态环境大数据数据字典的内容。

(6) 生态环境大数据数据字典内容属性：规定了生态环境大数据数据字典的内容属性、数据字典管理信息、数据库信息、其他关系数据库相关实体、用户定义实体、非结构化与半结构化数据信息、互联网数据资源信息以及区块链信息。

(7) 生态环境大数据数据字典文档编制要求：规定了生态环境大数据数据字典文档的编制要求、文档组织、文档存储格式与文档样例及编辑说明。

2. 适用范围

本文件规定了生态环境大数据数据字典的内容组成和属性填写要求，并提供常见的字典样式。

本文件适用于指导生态环境大数据设计开发单位进行系统分析和建模，也可作为生态环境大数据项目数据字典部分验收的参考依据。

3. 术语和定义

本部分为执行本标准制定的专门术语和对容易引起歧义的名词进行的定义。

(1) 数据项 data item

数据字典中具有独立含义的最小数据单位。

(2) 实体 entity

具体或抽象的事物及其关系。事物可以是人、物体、事件、观念、

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/205013022032011332>