

《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

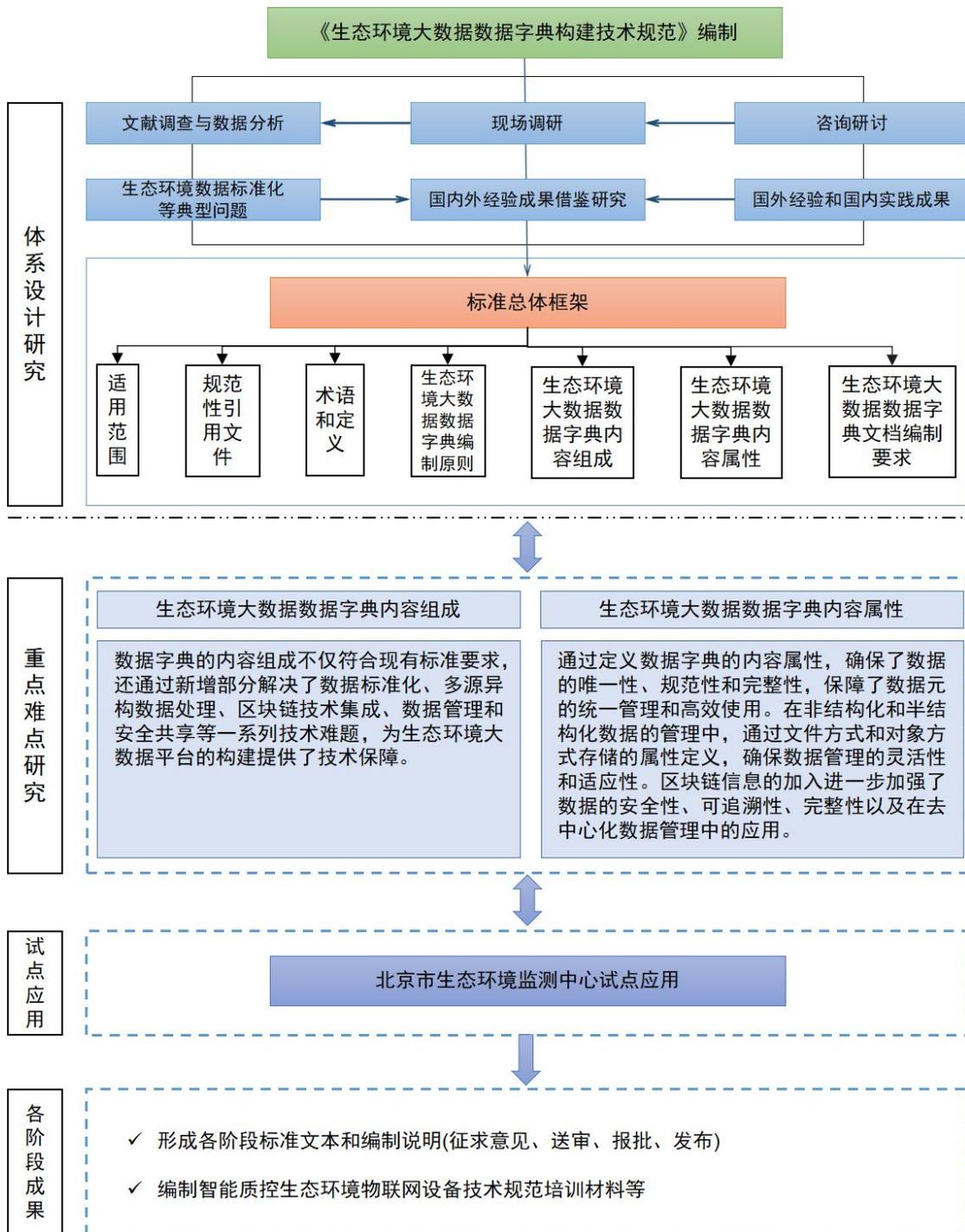
准的技术框架，明确了标准的主要内容、结构布局和编写要求。

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、

信息、过程等，实体的存在不依赖于是否有关于它的数据。

(3) 环境数据字典 data dictionary of environmental data

环境信息数据体系结构的描述信息的集合，记录了数据库的组成和格式等信息。环境信息数据字典属于环境信息元数据范畴，由实体和数据元组成。

(4) 统一资源标识符 uniform resource identifier, URI

统一资源标识符是一个用于标识某一互联网资源名称的字符串。该种标识允许用户对任何（包括本地和 互联网 ）的资源通过特定的协议进行交互操作。URI 由包括确定语法和相关协议的方案所定义。URI 的标准格式为：

[协议名]://[用户名]:[密码]@[服务器地址]:[服务器端口号]/[路径]?[查询字符串]#[片段 ID]

4. 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的编制应遵循以下原则。首先，编制过程应参考并兼容现有标准，确保数据项定义符合国家及环境行业的数据元、数据字典、数据分类和数据编码等相关标准，确保数据字典的规范性和依据性。其次，数据字典应结合实际应用，简洁实用，从系统开发和维护的角度出发，同时促进数据交换共享，确保内容的实用性和准确性。此外，数据字典应在系统需求和设计发生变更时及时更新，以保证数据字典与系统最终实现的一致性。最后，数据字典应具备开放兼容性和易于扩展的特点，遵循先标准后定义的建设规范，以保证数据定义的标准化和未来的扩展性。

5. 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的内容组成应符合 HJ/T 419 和 HJ 723 的相关要求，同时需满足数据安全可信和可控共享的要求。在编制数据字典时，应包括数据字典管理信息、数据库信息、数据表信息、视图信息、存储过程信息、用户函数信息、用户定义数据类型信息及数据项（字段）信息等基本内容。此外，鉴于生态环境数据的多样性，本标准还增加了区块链数据字典部分，涵盖区块信息、事务管理信息、实体信息、合约信息、账户信息和配置信息。

对于数据库信息，数据字典应包含数据源、端口、用户名、密码、编码方式、数据库名称等数据元，并参照数据库管理系统类型代码。数据表、视图、存储过程及用户函数信息需符合 HJ/723 标准，确保定义准确。此外，非结构化与半结构化数据的数据字典内容应包括文件存储和对象存储方式的数据元，例如文件名、格式、编码格式等。

对于互联网数据资源和区块链数据，字典内容应包含统一资源标识符和数据描述。区块链数据字典还包括账户、区块、事务、智能合约和配置等实体的详细数据元，如账户的公钥和私钥、区块的高度和时间戳、事务的标识和处理费用、合约的版本号及配置的协议版本号等。

6. 生态环境大数据数据字典内容属性

在生态环境大数据的环境信息数据字典中，内容属性包括数据的基本描述、管理要求、存储方式以及与区块链的集成。首先，数据字典内容属性涵盖了数据的名称、域名、是否为必填项、最大出现次数、

数据类型和值域。这些属性定义了每个数据元的标识和属性要求，确保数据的一致性和完整性。名称为数据元提供唯一标识，域名通常为缩略语，有助于简化数据元的引用。是否为必填项标明了数据元的必要性，而最大出现次数则规定了数据元的实例数量。数据类型描述了数据的格式，如整数、文本、日期等，值域则定义了数据允许的范围或自由文本。

数据字典管理信息则涉及数据字典的基本管理要求，如数据字典名称、版本、编写人及其日期等，确保数据字典的更新和维护信息完整透明。数据库信息属性进一步定义了数据库的各种配置，如数据源、端口、用户名、密码等，确保数据库连接的有效性和安全性。

对于非结构化与半结构化数据的存储，数据字典提供了文件方式和对象方式的详细属性要求，涵盖数据源、端口、用户名、密码等基本信息，并对文件名、编码格式、对象标识等进行具体说明，以支持对非结构化数据的有效管理。

在与区块链技术的集成方面，数据字典描述了区块链账户信息、区块信息、事务信息、智能合约信息和配置信息的属性要求。这包括账户的公钥、私钥及其资产，区块的标识、版本信息及难度系数，事务的标识、类型、金额等，以及智能合约的代码和存储。这些属性确保了区块链数据的准确性、可追溯性和安全性，同时支持区块链网络的高效运行和管理。

综上所述，数据字典的核心技术内容确保了环境信息数据的规范管理和应用，还通过详细的属性定义和技术要求，支持数据的有效存

储、管理和区块链集成。

7. 生态环境大数据数据字典文档编制要求

生态环境大数据平台的数据字典是系统分析阶段的关键工具，旨在详细描述数据元素、维护数据一致性并作为设计参考。编制数据字典文档时，需要遵循 GB/T 8567 的相关要求，确保文档的规范性和有效性。

数据字典文档应当单独组织，每个数据库实例需要形成一个独立的数据字典文档。这种组织方式确保了每个数据库实例的详细信息和管理要求可以得到清晰和系统的记录。文档应以电子版格式存储，推荐使用“.doc”、“.docx”、“.WPS”、“.xls”或“.xlsx”等通用格式，这些格式便于存储和共享。

在编制数据字典文档时，应参考附录 A 中的样例，并根据实际情况进行编辑。文档中的表格行数和列数不受限制，用户可以根据需要自行增加。附录中的斜体字仅作示例，实际编写时应替换为具体内容。用户可根据个人习惯和项目需求，自由设置字体、字号及段落格式。此外，编制过程中应根据项目的具体需求，灵活插入相应的内容，以确保数据字典的完整性和实用性。

六、标准实施的环境效益与经济技术分析

1. 环境效益分析

实施《生态环境大数据数据字典构建技术规范》将带来显著的环境效益。首先，通过统一的数据字典构建标准，可以极大地提升生态环境数据的一致性和准确性，减少因数据不一致导致的决策失误和资

源浪费。其次，标准化的数据字典有助于推动全国范围内生态环境数据的整合与共享，打破信息孤岛，实现跨地域、跨部门的数据互通，为政府决策、环境监管和生态保护提供更加全面、精准的数据支持。此外，标准化的数据字典还能够促进环保技术的创新和应用，推动生态环保领域的数据挖掘和分析工作，进一步提升生态环境保护的科学性和有效性。

2. 经济技术分析

从经济角度来看，虽然实施本标准需要一定的初期投入，包括标准的宣传培训、软件工具的开发和维护等费用，但这些投入将带来长期的经济回报。首先，标准化的数据字典能够降低数据管理和维护的成本，通过减少数据错误和重复工作，提高工作效率。其次，通过促进数据共享和整合，可以优化资源配置，避免重复建设和投资浪费。再次，标准化的数据有助于吸引更多企业和资本进入生态环保领域，推动相关产业的发展和壮大。最后，通过提高生态环境数据的利用价值，可以为政府和社会各界提供更加精准、高效的环保服务，促进经济社会的可持续发展。

此外，从技术角度来看，实施本标准也有助于提升我国生态环境大数据管理的技术水平。通过引入先进的数据字典构建技术和管理理念，可以推动我国生态环境大数据管理体系的现代化进程。同时，与国际标准接轨的数据字典构建标准也将有助于提升我国在国际生态环境大数据领域的竞争力和影响力。

七、标准实施建议

为确保《生态环境大数据数据字典构建技术规范》的有效实施，我们提出以下建议：

（一） 加强宣传培训：通过多种渠道和形式，广泛宣传本标准的重要性的意义，提高各级生态环境部门和相关单位对数据字典构建技术的认识和理解。同时，组织专题培训活动，邀请专家学者进行授课和指导，提升相关人员的专业水平和操作能力。

（二） 完善配套措施：结合本标准的实施要求，制定和完善相关配套措施和制度规定。例如，建立数据字典构建工作的责任机制和考核机制；制定数据字典的更新和维护流程；加强数据安全管理和隐私保护等。

（三） 强化监督检查：加强对本标准实施情况的监督检查力度，定期评估各地各单位的数据字典构建工作进展和成效。对于不符合标准要求单位和个人进行通报批评或采取相应措施予以纠正。

（四） 加强协作交流：鼓励各地各单位之间加强协作和交流，共同推动数据字典构建技术的创新和应用。通过建立信息共享平台或协作网络等形式，促进经验分享和技术交流，提高整体工作水平。

（五） 持续优化完善：根据实施过程中的反馈意见和建议，及时对本标准进行优化和完善。同时，关注国内外生态环境大数据管理领域的新技术和新方法，及时更新和补充标准内容，确保标准的先进性和适用性。