

《生态环境大数据平台数据接入技术规范》

（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。

数据接入是实现生态环境数据汇集的手段，是数据资源建设的基础环节，基础不牢，就无法为基于数据的应用提供稳定可靠的数据服务，不能充分发挥出数据的价值。不论业务数据接入各级信息中心还是各个业务部门之间的相互数据交换，都存在数据接入的问题。要实现不同所有者、不同权限要求、不同认证方式、不同业务类型、不同物理接口或介质、不同软件协议、不同数据库结构等多源多态数据的接入，保证数据接入过程的安全性、可靠性、稳定性和完整性，需要对数据接入的技术要求、流程、接口定义、接入数据质量等几个方面进行规范化和标准化。要充分发挥大数据技术在环境监测中的作用，实现智能、精准的环境监测监管，也需要建设相应的生态环境监测数据接入平台，制定相关数据接入标准规范，从各监测业务平台、外部

单位、社会互联网等获取丰富的数据源。

中华环保联合会积极响应国家号召，针对生态环境大数据接入技术规范的现实需求，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《生态环境大数据平台数据接入技术规范》团体标准的编制任务。

二、目的、意义、必要性

1. 目的

《生态环境大数据平台数据接入技术规范》旨在通过明确数据接入的范围、方式、质量管理、安全管理等方面的要求，为生态环境大数据平台的数据采集、存储、处理和应用提供一套科学、规范的技术指导。具体而言，该标准旨在解决数据接入过程中的异构性、不一致性、安全性等问题，确保数据的有效整合与高效利用。

2. 意义

促进数据共享与整合：在当前分部门、分领域的环境治理体系下，数据孤岛现象普遍存在。通过制定统一的数据接入标准，可以打破部门壁垒，实现跨部门、跨领域的数据共享与整合，为综合环境治理提供全面的数据支持。

提升数据质量：数据质量是环境治理决策的基础。通过明确数据接入的质量要求和技术规范，可以确保接入数据的准确性、完整性、时效性和一致性，为科学决策提供可靠依据。

加强数据安全与隐私保护：生态环境数据涉及国家安全和公众隐

私，其安全性不容忽视。通过引入区块链等先进技术，构建安全可信的数据接入与共享机制，可以有效保障数据在传输、存储、共享过程中的安全性和隐私性。

3. 必要性

在生态环境保护领域，大数据技术的应用日益广泛，为科学决策、精准施策提供了有力支持。然而，生态环境数据的多源异构性给数据整合、处理和分析带来了巨大挑战。为了有效应对这一挑战，制定“生态环境大数据平台数据接入规范”显得尤为必要：

(1) 确保数据的一致性与可比性

生态环境数据来源于众多不同的系统和设备，其数据格式、数据结构和数据质量千差万别。这种多源异构性导致数据在接入平台时容易出现不一致性和不可比性，严重影响后续的数据处理和分析结果。制定数据接入规范，可以明确数据接入的统一标准和格式，确保所有接入平台的数据都遵循相同的规则，从而提高数据的一致性和可比性。这不仅有助于提升数据处理的效率，还能为科学研究和政策制定提供更加准确、可靠的数据支持。

(2) 提升数据质量与可信度

数据质量是生态环境大数据平台运行效果的关键。低质量的数据不仅无法提供有价值的信息，还可能误导决策，造成不良后果。制定数据接入规范，可以设定严格的数据质量检查标准和流程，要求数据提供者在接入数据前进行必要的清洗、校验和修正。这些措施能够有效提升接入平台的数据质量，降低数据中的错误和异常，提高数

据的可信度和应用价值。

（3）促进数据共享与协作

生态环境保护是一项复杂的系统工程，需要多部门、多领域的共同参与和协作。数据共享是实现协同治理的重要基础。然而，由于数据格式、接口标准不统一等问题，数据共享往往面临诸多障碍。制定数据接入规范，可以为数据共享提供统一的标准和接口，降低数据交换的难度和成本。这有助于打破信息孤岛，促进生态环境数据的互联互通和共享利用，为跨部门、跨领域的协作提供有力支持。

（4）推动平台建设与标准化发展

生态环境大数据平台的建设是一个长期而复杂的过程，需要不断完善和优化。制定数据接入规范，可以为平台的建设提供明确的方向和指导，推动平台向标准化、规范化方向发展。这有助于提升平台的整体性能和稳定性，降低运维成本和维护难度。同时，标准化的数据接入规范还有助于吸引更多的数据提供者 and 使用者参与到平台建设中来，形成良性发展的生态系统。

综上所述，制定“生态环境大数据平台数据接入规范”是确保数据一致性与可比性、提升数据质量与可信度、促进数据共享与协作以及推动平台建设与标准化发展的必然要求。该标准的制定将为生态环境大数据平台的建设和运行提供有力保障，为生态环境保护事业注入新的活力。

三、主要工作过程及标准编制技术路线

1. 主要工作过程

本标准由北京环丁环保大数据研究院牵头组织编制，并联合了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单位共同参与。编制工作自启动以来，经历了以下几个关键阶段：

前期调研：编制组对国内外相关标准、技术规范、政策法规进行了广泛调研，收集了大量的背景资料和技术参考。

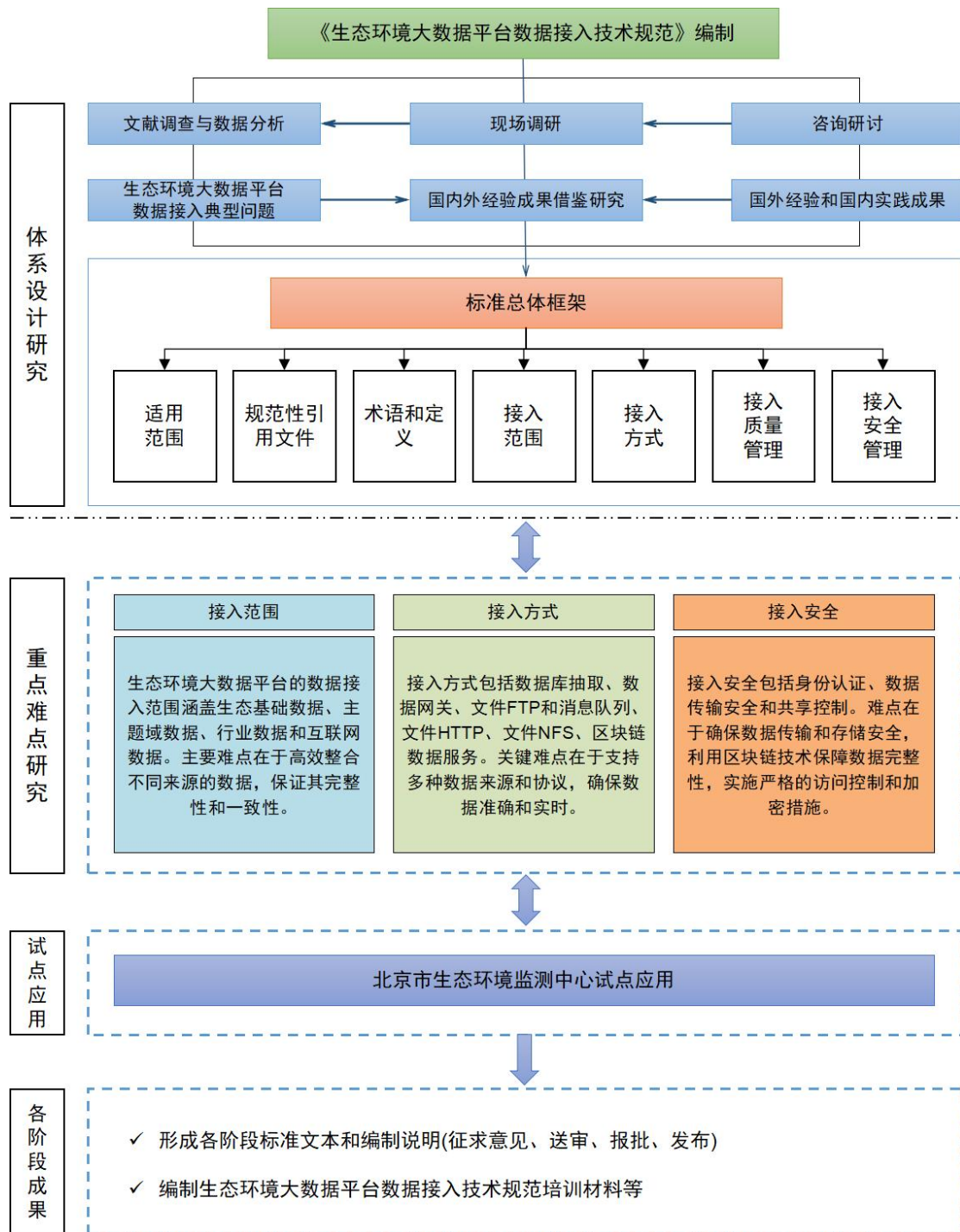
框架设计：在充分调研的基础上，编制组确定了标准的总体框架和主要内容，包括接入范围、接入方式、接入质量管理、接入安全管理等核心章节。

草案编制：根据框架设计，编制组分工协作，完成了标准草案的初稿编写工作。初稿内容涵盖了标准的适用范围、规范性引用文件、术语和定义、接入范围、接入方式、接入质量管理、接入安全管理等各个方面。

内部评审：初稿完成后，编制组内部进行了多次评审会议，对草案内容进行了逐条讨论和修改完善。同时，还邀请了相关领域的专家进行内部咨询和指导。

召开立项会：2024年6月5日，标准立项会召开。与会专家通过质询和讨论，评估了拟立项标准的必要性和可行性，一致认为该项目符合立项条件。

2. 标准编制技术路线



四、国内外相关标准研究

1. 国内同类标准状况

在我国，随着大数据技术的快速发展和广泛应用，生态环境大数

据平台的建设和运营日益受到重视。为了规范平台的数据接入过程和提高数据质量水平，国内已经出台了一系列相关标准和技术规范。例如 DB15/T 1873-2020《大数据平台数据接入质量规范》、DB35/T 1893-2020《生态环境大数据管理平台接口规范》等地方标准以及 DB31/T 1240.2-2020《平台接入技术要求》等行业标准均为生态环境大数据平台的数据接入提供了技术指导和规范依据。这些标准在数据类型定义、接口设计原则、数据质量评估等方面均有所涉及但各有侧重。相比之下，《生态环境大数据平台数据接入技术规范》在综合考虑国内外先进经验和实际需求的基础上更加注重全面性和系统性，旨在为我国生态环境大数据平台的数据接入提供一套完整、科学的解决方案。

2. 国际同类标准状况

在国际上，大数据技术在生态环境治理中的应用也日益广泛且深入。虽然直接针对生态环境大数据平台数据接入的技术规范相对较少但大数据技术在各行业领域的接入和管理方面已经形成了较为成熟的标准体系和实践经验。例如 ISO/IEC 等国际标准化组织以及各国标准化机构均发布了大量与大数据处理和分析相关的国际标准和规范这些标准在数据格式、接口协议、安全保护等方面提供了有益的借鉴和参考。如 2020 年国际电工委员会发布的 Internet of Things (IoT)—Data exchange platform for IoT services — Part 1: General requirements and architecture (ISO/IEC 30161-1:2020)。标准为了建立一个共同的平台，以支持日益多样化

的 IoT 用例和应用程序,规定了物联网 (IoT) 数据交换平台 (DEP) 的一般要求和架构,该平台支持各种 IoT 服务与传统服务的共存和高效部署。IoT DEP 提供中间件组件,以促进 IoT 和传统服务之间的通信,并支持不同 IoT 服务之间的数据交换。然而需要注意的是由于国内外在生态环境治理体系、政策法规、技术环境等方面存在差异因此国际标准的直接应用可能存在一定的局限性和挑战。《生态环境大数据平台数据接入技术规范》在编制过程中充分借鉴了国际先进经验并结合我国实际情况进行了适当的创新和调整以确保标准的适用性和有效性。

五、主要技术内容及说明

1. 标准框架

本标准主要包括适用范围、规范性引用文件、术语和定义、接入范围、接入方式、接入质量管理、接入安全管理等方面内容。

(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) 接入范围:规定了智能质控生态环境物联网设备可以接入的各类数据资源的范围,包括生态环境基础信息、主题域数据、行业相关数据和互联网数据。

(5) 接入方式:规定了智能质控生态环境物联网设备接入生态环境大数据平台的多种方式,包括关系数据库抽取、数据网关、文件接收与拉取 FTP 服务、消息队列、文件 HTTP 服务、文件 NFS 服务,

以及区块链数据服务等，确保数据的完整性、传输的安全性以及多种接入方式的兼容性。

(6) 接入质量：规定了数据质量评价维度、数据接入质量技术要求和数据质量评分方法。

(7) 接入安全：规定了身份认证、数据传输和数据共享需基于区块链技术，以确保安全性。

2. 适用范围

本标准规定了生态环境大数据平台的接入范围、接入方式、接入质量管理、接入安全管理。

本标准适用于生态环境大数据平台开发单位进行数据采集服务功能研发、数据采集工具选型及其他数据接入场景提供规范要求。

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) 大数据 big data

具有体量巨大、来源多样、流动高速、形态多变等特征，并难以用传统数据体系结构有效处理的数据集。

(2) 数据接口 data interface

预先定义的为应用程序提供实现数据访问、数据服务和数据服务等功能的协议、程序、函数或过程。

(3) 结构化数据 structured data

数据结构固定，能够用统一的结构表示的数据。本规范指存储于

业务系统关系数据库中的数据。

(4) 非结构化数据 unstructured data

数据结构不固定，且不能够用统一的结构表示的数据。包括所有格式的办公文档、文本、图片、XML、HTML、各类报表、图像、音频和视频信息等数据。

4. 接入范围

接入范围主要依据生态环境大数据平台的实际需求来确定，旨在明确哪些数据应当被纳入平台进行管理和分析。

本部分编写内容参考了多个标准文件，如 HJ 719-2014、HJ 721-2014、HJ 727-2014、DB15/T 1873-2020 等，以及实际业务需求，来界定数据接入的范围。

生态环境大数据平台的数据接入范围旨在支持全面的环境数据整合与分析，为环境监测和管理提供坚实的数据基础。生态环境基础信息资源包括环境质量、污染源、环境管理及其他四个一级类目。在环境质量方面，涵盖水环境质量、大气环境质量、土壤环境质量、自然生态环境质量、噪声环境质量和海洋生态环境质量等。污染源方面包括废水、废气、固体废物、重金属、挥发性有机物等多种类型，以及环境应急风险源和交通运输污染源等。环境管理领域则涉及污染物排放控制、生态保护、气候变化应对、固体废物与化学品管理、环境安全应急管理、环境行政许可、环境监测及环境保护督察等方面。其他类目包括环境行政办公、综合管理、科技、政策法规、标准规范等。

主题域数据资源涵盖了水环境、大气环境、土壤环境、海洋生态

环境、自然生态、固废与危化品、核与辐射、噪声、气候变化应对、行政许可、生态环境监测、环境督察、环境执法、环境统计、环境应急、环境舆情管理等领域。这些数据支持平台对生态环境的多维度分析与监控。

行业相关数据资源则包括来自其他管理部门的数据，例如住房和城乡建设部门提供的工地和污水处理厂数据、公安部门的机动车数据、发展和改革委员会的燃煤数据及项目建设数据、水利部门的水文数据、自然资源部门的地下水和矿权数据、农业农村部门的土壤数据、气象部门的大气数据、海洋部门的功能区划和入海水源数据等，这些数据帮助实现跨部门环境信息的整合。

互联网数据资源包括来自生态环境相关的各种网络数据，如政府网站、系统网站、媒体网站、科技期刊、社交媒体等，为平台提供了广泛的网络信息支持。

5. 接入方式

接入方式部分内容的确定主要基于生态环境大数据平台的技术架构、数据源的多样性以及数据处理的需求。

本部分内容参考了诸如 GB/T 35274-2017、GB/T 36345-2018 等标准文件，同时考虑了实际应用场景和技术可行性。

(1) 关系型数据库抽取

在生态环境大数据平台的数据接入方式中，关系型数据库抽取的核心技术内容涉及到多个关键功能和操作要求。关系型数据库抽取的主要目的是将数据从关系型数据库迁移到大数据平台，确保数据的完

整性和可用性。系统必须支持对主流关系型数据库进行数据提取，包括数值型、字符型、日期/时间型等常见数据类型。此外，抽取过程应提供“全量”和“增量”两种模式，以满足不同的数据同步需求。

在技术实现方面，系统需要能够从关系型数据库中抽取结构化数据和半结构化数据，支持数据表、字段选择和字段类型转换等操作。数据抽取操作应支持立即执行或按照定期、不定期的调度策略进行，调度策略应包括固定间隔、指定时间点和范围等多种选项。为提升用户体验，系统应提供图形化管理界面，方便配置抽取模式、源数据库、目标存储位置等，同时具备完善的日志记录和审计功能，以记录操作配置和运行时事件。此外，系统应具备监控机制，以便在运行过程中快速定位和解决异常问题。

在应用场景中，关系型数据库抽取服务涉及从数据源中提取数据库的元数据信息，配置数据传输任务，并定义任务的运行策略。这包括设定数据源表和大数据平台的目标表，配置任务的执行时间、频率等参数，并进行实际的数据传输操作，将数据从源数据库抽取到目标平台存储中。这些技术和操作要求确保了数据在生态环境大数据平台中的有效集成和管理。

(2) 数据网关

在生态环境大数据平台的数据接入方式中，数据网关的核心技术内容主要围绕提供灵活的接口数据接入服务。数据网关应具备对各种数据接口的支持能力，包括 Webservice 和 RESTful 接口，以适应结构化和非结构化数据的接入需求。通过接口编排功能，数据网关可以

实现多个接口的集成，使得不同数据源的数据能够高效整合。

数据网关服务还包括图形化管理界面，这一界面用于配置接口数据的存储位置、操作用户和目标存储位置，提升了数据管理的便捷性。系统应具备完善的日志记录和审计功能，能够详细记录接口数据的配置及数据抽取操作的各类事件。此外，为保证服务的稳定性和可用性，数据网关必须具备熔断管理机制，以应对接口访问异常情况，确保服务的整体可靠性。

在应用场景中，数据网关服务处理数据源向大数据平台的数据传输任务。数据源提供接口信息，包括接口访问地址、输入输出参数和验证方式等，大数据平台则根据这些信息进行接口定义和编排。数据网关服务配置数据传输任务的运行策略，包括任务的开始时间、结束时间和运行频度等，并负责从数据源的数据接口中抽取数据，将其传输到大数据平台的数据仓库。这些功能确保了数据的高效接入和管理，为平台的运行提供了坚实的支持。

（3）文件接收 FTP 服务

文件接收 FTP 服务主要用于实现外部系统文件数据的接收，并将这些文件存储到大数据平台的数据仓库中。该服务的核心技术内容包括支持标准 FTP 协议，以确保数据传输的兼容性和可靠性。它还具备顺序型断点续传功能，这对于处理因网络中断或其他问题导致的文件传输中断至关重要，确保文件能够从中断处继续传输而无需重新开始。此外，服务支持文件的重命名及存储目录的指定，便于对接收到的文件进行管理和分类。

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/186012002025010232>