

DB 3201

南 京 市 地 方 标 准

DB 3201/T XXXX—XXXX

城市生态环境 数据管理通用要求

Urban ecological environment—General requirements of data management

(报批稿)

XXXX—XX—XX 发布

XXXX—XX—XX 实施

南京市市场监督管理局 发 布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语、定义和缩略语 1

 3.1 术语和定义 1

 3.2 缩略语 2

4 数据管理架构 2

 4.1 管理原则 2

 4.2 管理架构 2

5 数据生存周期管理要求 3

 5.1 数据采集 3

 5.2 数据存储 5

 5.3 数据处理 5

 5.4 数据共享 6

6 数据治理要求 7

 6.1 数据目录管理 7

 6.2 元数据管理 7

 6.3 数据追溯 8

7 数据质量管理要求 8

 7.1 基本要求 8

 7.2 数据质量要求 8

 7.3 质量控制要求 9

 7.4 数据质量检查方式与方法 10

8 数据安全要求 10

 8.1 安全要求 10

 8.2 备份要求 11

 8.3 恢复要求 11

参考文献 12

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由南京市生态环境局提出并归口。

本文件起草单位：中电莱斯信息系统有限公司、南京市生态环境保护科学研究院、中国科学院生态环境研究中心、南京市生态环境信息服务中心、中国标准化研究院。

本文件主要起草人：彭向阳、周艳文、王辰星、隋媛、鲁士仿、王骊、郝元、周解臻、严岩、王泽童。

城市生态环境 数据管理通用要求

1 范围

本文件规定了城市生态环境数据管理架构、数据生存周期管理要求、数据治理要求、数据质量管理要求和数据安全要求。

本文件适用于城市生态环境数据的管理和运维。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB/T 13989 国家基本比例尺地形图分幅和编号
- GB/T 17278 数字地形图产品基本要求
- GB 17859 计算机信息系统 安全保护等级划分准则
- GB/T 20273 信息安全技术 数据库管理系统安全技术要求
- GB/T 22239 信息安全技术 网络安全等级保护基本要求
- GB/T 22240 信息安全技术 网络安全等级保护定级指南
- GB/T 37025 信息安全技术 物联网数据传输安全
- GB/T 39477 信息安全技术 政务信息共享 数据安全技术要求
- HJ/T 419 环境数据库设计与运行管理规范
- HJ 719 环境信息系统数据库访问接口规范
- HJ 723 环境信息数据字典规范
- HJ 724 环境基础空间数据加工处理技术规范
- HJ 726 环境空间数据交换技术规范
- HJ 727 环境信息交换技术规范
- HJ 966 生态环境信息基本数据集编制规范
- DB3201/T ×××× 城市生态环境 信息资源目录建设要求

3 术语、定义和缩略语

3.1 术语和定义

HJ 724界定的以及下列术语和定义适用于本文件。

3.1.1

数据管理 data management

数据资源获取、控制、价值提升等活动的集合。

[来源：GB/T 34960.5—2018，定义3.2]

3.1.2

空间数据 spatial data

用来表示空间实体的位置、形状、大小和分布特征诸方面信息的数据，适用于描述所有呈二维和多维分布的关于区域的现象。

[来源：HJ 724—2014，定义3.10]

3.2 缩略语

下列缩略语适用于本文件。

API：应用程序编程接口（Application Programming Interface）

4 数据管理架构

4.1 管理原则

4.1.1 在数据生存全周期内，开展数据质量管理和数据安全管理工作。

4.1.2 根据数据来源、数据共享限制等对数据进行分类和分级管理，数据分类和分级符合相关制度和标准的要求。

4.1.3 参与城市生态环境数据相关工作的组织、个人都是数据管理的主体，确保数据的客观真实。

4.2 管理架构

生态城市环境数据管理架构见图1，包括

- a) 数据生存周期管理：
 - 1) 数据采集；
 - 2) 数据存储
 - 3) 数据处理；
 - 4) 数据交换与共享；
- b) 数据治理：
 - 1) 数据目录管理；
 - 2) 元数据管理；
 - 3) 数据追溯；
- c) 数据质量管理；
- d) 数据安全管理工作。

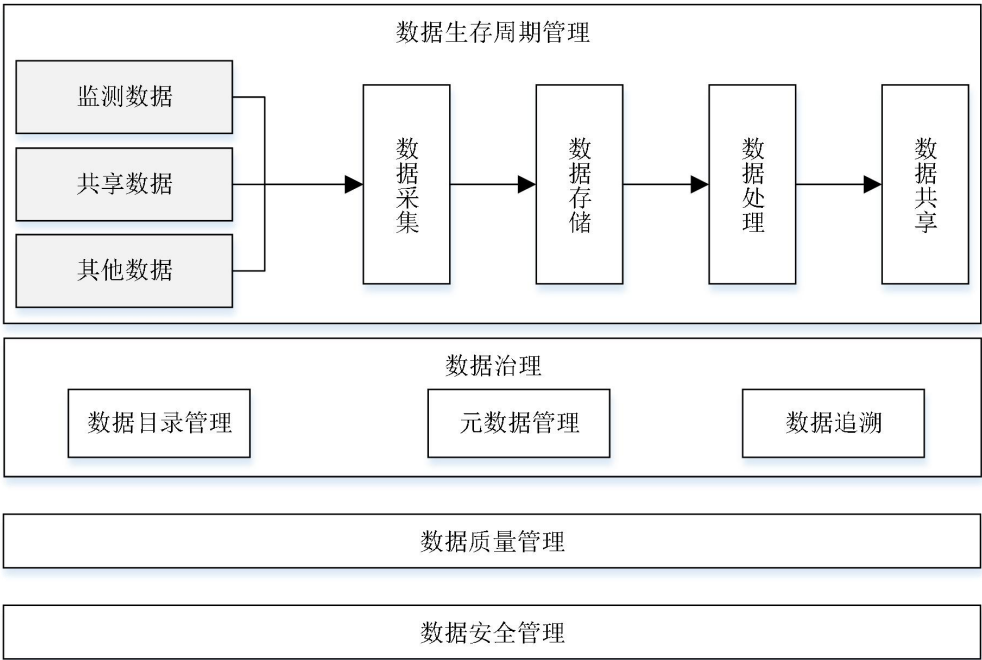


图 1 城市生态环境数据管理架构

5 数据生存周期管理要求

5.1 数据采集

5.1.1 数据源

5.1.1.1 自动监测数据

主要包括自动监测站点实时监测数据，由自动采集设备进行实时采集或人工不定时监测获得。经由采集设备前端数据库传输到各相关数据负责单位或数据中心，通过共享设置可共享交换到其他使用单位。

对于实时监测数据应尽量避免人工干预。监测站点宜根据数据实际情况确定数据值的有效范围和预警范围，便于自动判断数据的有效性并进行数据预警。

5.1.1.2 外单位共享数据

包括上位系统的回传数据及其他单位的共享交换数据。

5.1.1.3 其他数据

其他途径获得的数据。

5.1.2 数据内容

本文件中城市生态环境数据内容参见DB3201/T ××××《城市生态环境 信息资源目录建设要求》，包括但不限于：

- a) 生态环境质量数据；

- b) 自然生态环境数据;
- c) 污染源数据;
- d) 生态环境管理业务数据;
- e) 生态环境科技及其管理数据;
- f) 生态环境保护产业数据;
- g) 环境政策法规标准;
- h) 环境保护相关数据;
- i) 其他生态环境数据。

5.1.3 数据类型

城市生态环境数据按数据结构可分为:

- a) 结构化数据: 电子化、规范化的数据, 数据项设计、格式等均满足融合需求, 无需做进一步的处理即可使用, 主要包括数据库数据、电子表格数据、地理信息数据等;
- b) 半结构化数据: 电子化、规范化的数据, 但数据项格式需进行一定处理后, 才能满足数据融合的使用要求, 主要包括文本、网页、日志文件、XML 文件等;
- c) 非结构化数据: 没有明确结构约束的数据, 主要包括图片、图像、音频、视频、纸质扫描件等。

5.1.4 采集方法

城市生态环境数据接入方法包括但不限于:

- a) 数据库: 建立源数据库结构与目标数据库之间的存储结构映射, 通过数据库同步进行采集, 数据库建设与运维符合 HJ/T 419 的要求;
- b) 文件: 通过文件导入进行采集, 对于非结构化文件数据, 可将文件整体上传至服务器后建立存储路径表;
- c) API 接口:
 - 1) 针对结构化和半结构化文件, 应建立源接口返回数据结构与目标数据库之间存储结构映射, 可通过搜索进行接口数据采集, 也可直接存储接口的基本信息;
 - 2) 针对返回非结构化接口类型, 直接存储接口基本信息;
- d) 消息队列: 建立源数据结构与目标数据库之间存储结构映射, 通过订阅消息队列进行采集。

5.1.5 采集要求

数据采集应符合的要求包括但不限于:

- a) 明确采集数据目的和用途, 明确采集源、采集范围、采集频率;
- b) 开展全面分析, 设置数据采集策略, 如采集周期、采集方式、采集内容等:
 - 1) 业务分析, 对数据来源业务进行分析;
 - 2) 接入方式分析, 对源数据存储位置、接入方式进行分析;
 - 3) 结构分析, 对数据的含义、类型、长度、结构进行分析;
 - 4) 内容分析, 对数据长度、分布状态、平均值、中位数、极值等进行统计分析;
 - 5) 关联分析, 对数据之间存在的依赖关系、主外键关系进行分析;
- c) 制定采集数据的清洗、转换、加载等操作规范, 做好采集数据的备份工作;
- d) 对不同类别、级别的数据制定并实施不同的采集策略, 做好采集过程的安全防护措施;
- e) 遵循数据最小化原则, 只采集满足业务所需的最少数据;
- f) 制定采集数据的质量保障规则, 明确数据质量保障的策略、规程和要求;

g) 对采集行为进行日志记录和安全审计。

5.2 数据存储

5.2.1 传输要求

数据传输应符合以下要求：

- a) 依据数据的分类分级，明确数据传输的场景，建立相应的数据传输策略；
- b) 采取必要的措施保障传输通道、传输节点和传输数据的安全、准确和及时；
- c) 建立传输数据完整性检测和数据恢复控制措施。

5.2.2 存储要求

数据存储应符合以下要求：

- a) 对不同类别、不同级别的数据采用差异化安全存储，选择适合的加密算法、去标识化等手段进行加密存储；
- b) 支持对结构化数据及非结构化数据的存储能力：
 - 1) 对结构化数据，选择结构化数据库存储；
 - 2) 对半结构化数据，转为结构化数据后选择结构化存储；
 - 3) 对非结构化数据，根据存储量选择合适的文件系统进行存储；
- c) 建立数据的容错、高可用、备份和快速恢复机制；
- d) 存储设备性能可靠，业务数据存储系统应留有一定容量；
- e) 宜优化数据组织方式，按照用途、用户、权限等维度对数据封装打包。

5.3 数据处理

5.3.1 清洗比对

数据清洗对比内容包括但不限于：

- a) 残缺数据处理：确定范围、去除重要性低字段、填充缺失内容；
- b) 错误数据处理：处理格式内容错误、逻辑错误、不合规等问题；
- c) 重复数据处理：对重复出现的数据进行整合。

5.3.2 标签管理

以每条记录为单位进行标签处理，标签将跟随该数据记录在后续数据清洗、加工、整合等过程中流转，实现数据的溯源。

标签应具有唯一性。

5.3.3 数据校验

校验结果应准确、清晰、明确、客观，并进行唯一标识。数据校验方法包括但不限于：

- a) 数据可视化：通过数据可视化工具，将数据转换为图形或图表的形式，更直观地理解和分析数据；
- b) 数据预处理：通过多种预处理算法对数据进行预处理，如空值插补、去重、字段过滤等；
- c) 数据比较：将数据与预期结果进行比较，以确定数据是否存在问题；
- d) 数据统计分析：通过数据统计分析方法，如平均值、中位数、标准差等，对数据进行描述和分析，进一步发现异常值和趋势等问题；
- e) 数据建模：通过建立数据模型，对数据进行预测和分类，验证数据的准确性和有效性；

- f) 数据审计：对数据的来源、传输、存储等环节进行审计，确保数据的完整性、保密性和可用性。

5.3.4 关联映射

关联映射方式包括但不限于：

- a) 目标结构分析：根据源数据结构和目标结构的模型进行比对，针对字段名等进行分析比对，分析数据项映射关系，实现源字段和目标字段映射的自动匹配；
- b) 事实表比对映射：根据目标结构分析和数据转换规则，实现源数据结构的类型转换、字段拆分、字段合并、字符串处理、日期转换、算术运算转换等数据转换；
- c) 代码比对映射：根据代码数据转换规则完成源数据代码与标准代码比对映射。

5.3.5 数据整合

数据整合方式包括：

- a) 水平整合：
 - 1) 同一实体相同维度的数据进行水平整合，数据间存在结构差异的进行统一；
 - 2) 对不同来源的冲突数据，能判别数据有效性、正确性则保留符合项，否则应追溯增加数据来源标识后保留。
- b) 垂直整合：
 - 1) 对同一实体不同维度的数据进行垂直整合；
 - 2) 识别并提取出有效的业务主键，根据业务主键进行关联整合；
 - 3) 保留整合数据的来源信息；
 - 4) 对字段重合度低的数据，采用主从表的方式进行整合。

5.3.6 数据挖掘

- a) 通过统计分析类算法对数据进行统计分析，如协方差矩阵、方差、标准差等；
- b) 使用机器学习算法执行机器学习任务，如线性回归、决策树等，算法类型包括但不限于文本分析、分类、聚类、回归、推荐、关联分析等；
- c) 采用大数据、人工智能等先进技术，对城市生态环境数据进行统计和分析，形成多维度的数据报表为业务决策提供支撑。

5.4 数据共享

5.4.1 数据接口

5.4.1.1 数据库接口

数据库交换接口应符合HJ 719的要求，并具有以下功能：

- a) 数据库交换请求消息文件可直接放到指定目录下由数据交换平台获取，或通过 API 调用方式将请求消息直接发送到数据交换平台；
- b) 数据库交换消息应包含一个发送方、一个数据源及数据查询条件、多个接收方及数据接收目的；
- c) 消息接口应提供通用语言接口，满足不同系统的应用需求。

5.4.1.2 文件交换接口

文件交换接口应具有以下功能：

- a) 文件交换请求消息文件可直接放到指定目录下由数据交换平台获取，或者通过 API 调用方式将请求消息直接发送到数据交换平台；
- b) 文件交换请求消息应描述多个被发送文件，包括文件名、文件目录和传输次序等信息。
- c) 文件能够传输到多个接收方，在消息文件中描述多个接收方代码信息；
- d) 消息接口应提供通用语言接口，满足不同系统的应用需求。

5.4.2 数据交换要求

环境信息交换应符合HJ 727的要求，环境空间数据交换应符合HJ 726的要求。

城市生态环境数据交换还应符合以下要求：

- a) 支持对外交换的数据目录管理功能；
- b) 支持多用户并发调用；
- c) 对不同类别、级别的数据制定和实施不同的交换策略和交换过程的安全防护措施；
- d) 建立明确的数据开放和共享场景，确保数据的使用权限和使用范围；
- e) 建立数据共享访问控制机制，明确不同机构或部门、不同身份与目的的用户权限；
- f) 对共享数据的使用者提出明确的数据安全防护要求，在共享数据前对使用者进行数据安全风险评估；
- g) 建立数据共享审批流程，明确共享数据内容、交接方式及应用范围等；
- h) 建立数据公开发布审批制度，明确数据公开的内容及范围；
- i) 在数据公布之前，对拟公布数据的敏感性进行评估，根据评估结果对需要公布的敏感信息进行脱敏操作；
- j) 建立数据接口控制策略，明确数据接口安全要求、控制措施，包括接口名称、接口参数等；
- k) 与数据接口调用方明确数据的使用目的、供应方式、保密约定、数据安全责任等。

6 数据治理要求

6.1 数据目录管理

城市生态环境信息资源目录应符合DB3201/T XXXX《城市生态环境 信息资源目录建设要求》的要求。

数据目录管理还应符合以下要求：

- a) 满足多元、异构、海量信息资源的管理需要；
- b) 支撑用户多途径、多视角检索；
- c) 针对不同用户、业务目录，提供不同权限管理。

6.2 元数据管理

元数据管理应符合以下要求：

- a) 建立元数据管理机制，明确元数据的管理过程及角色职责；
- b) 明确元数据管理的范围和优先级，构建元数据库；
- c) 建立完整的数据字典、数据模型、数据架构及其管理体系，数据字典和数据集符合 HJ 723 和 HJ 966 的要求；
- d) 建立元数据创建、维护、整合、存储、分发、查询、报告和分析机制；

- e) 建立元数据管理的质量标准和评估指标，开展元数据绩效评估并持续改进。

6.3 数据追溯

应建立城市生态环境数据可追溯机制，如通过审计日志等方式在采集、审核校验、更正等流程中进行唯一性标识，并对相关记录和文件进行归档。

7 数据质量管理要求

7.1 基本要求

应建立城市生态环境数据质量管理体系及实施机制，优化数据质量并持续改进，满足数据应用的需求：

- a) 建立数据质量管理机构和机制，明确数据质量管理的角色和职责，建立数据质量管理方法；
- b) 明确不同数据之间的关系和依赖性，制定数据质量管理目标；
- c) 识别数据生存周期各个阶段的数据质量元素，构建数据质量评估框架；
- d) 采用定性评估、定量评估或综合评估等方法，评估和持续优化数据质量；
- e) 建立数据管理质量管理记录。

7.2 数据质量要求

7.2.1 文档表格数据

文档中信息要与真实情况相符，文档数据质量应符合以下要求：

- a) 文档、表格数据无缺失；
- b) 文档内容无缺漏，数据记录、统计无误，逻辑清晰；
- c) 文档结构、命名符合相关标准、规范的要求；
- d) 表格中数据定性和定量数据与其真值一致，计算结果正确。

7.2.2 空间制图数据

空间制图符合以下要求：

- a) 根据制图区域和制图对象的特点和分布规律，设置主题信息的分类、分级及符号化方法，选择合适的波段组合以及拉伸方式，按照投影坐标系要求，保证主图内容醒目、突出；
- b) 图名、图例、图廓、经纬网、比例尺、指北针、文字说明等制图要素应保持符号风格、颜色、粗细等协调，简单精炼；
- c) 出图分辨率满足业务需求，宜不低于 300 dpi。

7.2.3 空间数学基础数据

数学基础数据符合以下要求：

- a) 平面坐标系采用“2000 国家大地坐标系（CGS2000）”；
- b) 高程基准采用“1985 国家高程基准”，高程范围的度量单位描述采用米（m）；
- c) 投影方式按照 GB/T 17278 执行；
- d) 按照行政区域组织的数据可不分带，采用地理坐标，经纬度值采用“°”为单位，双精度浮点数表示；
- e) 比例尺标准分幅及编号按 GB/T 13989 执行。

栅格数据应创建金字塔结构。栅格数据转矢量时应满足 HJ 726 要求。

7.2.4 遥感影像数据

遥感影像数据应符合以下要求：

- a) 数据完整：数据名称、内容和格式等完整；
- b) 资料完备：数字影像数据、遥感器参数（遥感器类型、扫描带宽、空间分辨率、光谱数据资料完备性 分辨率以及其他特征参数）、遥感数据获取时的参数（太阳高度角、轨道高度、太阳倾角、重复周期等）等资料齐全；
- c) 信息直观效果：目标地物的大小、形状、阴影、色调、纹理等解译标志信息突出、明显；
- d) 影响重叠率：相邻景影像之间的重叠率在 4 %以上，云量较多、数据质量较差等特殊情况下不少于 2 %；
- e) 视角选择：山区、平地地区的影像侧视角分别不大于 16° 和 20° ；
- f) 辐射校正：精度达到 85%以上，光谱信息丰富，目标物清晰可见；
- g) 几何校正：对于平原和丘陵、山地，中分辨率和中分辨率影像控制点残差应分别 ≤ 1 倍、 ≤ 2 倍影像分辨率，高分辨率影像控制点残差应分别 ≤ 1 倍、 ≤ 3 倍影像分辨率；
- h) 影像配准：平原配准精度为 0.5 个像元，山区配准精度为 1 个像元；
- i) 影像镶嵌：影像镶嵌前接边误差小于 1 倍像素分辨率，镶嵌后影像清晰、色彩均匀；
- j) 影像融合：目视无重影、无模糊及光谱失真现象，边界清晰、无明显错位，能够反映细部特征；
- k) 校正精度：对于平原和丘陵、山地，中分辨率影像校正精度分别 ≤ 2 倍、 ≤ 3 倍影像分辨率，中分辨率影像校正精度分别 ≤ 2 倍、 ≤ 5 倍影像分辨率，高分辨率影像校正精度分别 ≤ 2 倍、 ≤ 10 倍影像分辨率。

7.3 质量控制要求

7.3.1 数据质量审核

城市生态环境数据质量审核要求包括但不限于：

- a) 建立数据质量审核程序，对于城市生态环境信息化系统录入和产生的数据的完整性、一致性等方面进行审核；
- b) 根据数据质量管理目标和计划对所录入的相关数据的质量进行自动或人工检验；
- c) 定期对数据质量进行核实确认；
- d) 对数据质量审核的证实性文件和记录进行归档。

7.3.2 数据质量标记

城市生态环境数据质量标记要求包括但不限于：

- a) 建立数据质量标记的业务规则，并负责监督、指导相关部门及业务人员对业务规则进行执行；
- b) 对数据质量进行统一标记，其他人员识别数据质量问题确认后进行标记。

7.3.3 数据更正

城市生态环境数据更正要求包括但不限于：

- a) 建立数据更正的业务规则，并负责监督、指导相关部门及业务人员对业务规则进行执行；
- b) 在审核校验过程中，若发现质量不合格的数据，应要求采集人员重新核对，与相关数据来源部门确认后对原录入数据进行更正；
- c) 对已标记的数据，应及时识别，并按相应程序对标记数据实施更正，对标记数据更正后，应撤销标记；

- d) 对其他方式发现的数据质量问题，应督促更正来自其他方式发现的质量不合格数据；
- e) 对数据更正记录进行归档。

7.4 数据质量检查方式与方法

7.4.1 检查方式

城市生态环境数据质量检查的方式包括：

- a) 全部检查：对批数据中的所有个体进行逐一检查，采用计算机自动检查方法时应进行全部检查。
- b) 抽样检查：按照确定的抽样方案，采用适当的抽样方法从批数据中抽取部分个体作为样本进行检查，并根据样本质量来判断批数据质量。进行抽样检查的数据应是在统一的标准或规范指导下，由基本相同的数据源、在基本相同的时段和软硬件条件下生产的数据。

7.4.2 检查方法

数据质量检查方法包括但不限于：

- a) 自动检查：使用数据质量检查软件实现信息系统自动检查，如计算检查、数据模板分析检查、空间约束分析检查；
- b) 人工检查：检查人员根据有关标准或要求、专家知识及经验进行的数据检查，如资料对比检查、经验判断检查等；
- c) 人机交互检查：在不能实现计算机自动检查时，可先通过计算机程序检索出疑问点，再采用人机交互进行检查，最后进行人工判断，如放大漫游检查、图层叠加检查、数据关联检查和数据格式转换检查。

8 数据安全要求

8.1 安全要求

城市生态环境数据安全应符合GB 17859、GB/T 22239、GB/T 22240、GB/T 37025、GB/T 39477等对数据安全的要求。

城市生态环境数据安全还应符合以下要求：

- a) 数据库建设应符合 GB/T 20273 的要求，能够提供字段级、记录级以及文件级的加密与完整性保护功能；
- b) 数据访问权限控制：
 - 1) 对数据资源设置不同的访问权限，支持基于角色的访问控制；
 - 2) 对用户进行分组管理，并基于用户组进行授权；
 - 3) 利用安全协议与证书对用户进行认证；
- c) 数据加密：
 - 1) 对于特定的敏感字段或业务数据使用加密存储；
 - 2) 根据数据的保密要求限定访问源，敏感数据应只接受来自内网、专线或虚拟专网的访问请求，并通过加密或其他有效措施实现传输保密性；
 - 3) 使用国家密码主管部门认证核准的密码产品，遵循密码相关的国家标准和行业标准；
 - 4) 按照国家相关保密部门要求选择加密算法；
- d) 安全审计：
 - 1) 通过日志记录用户对资源的访问行为；

- 2) 开展安全审计，审计范围应覆盖业务数据的全部用户行为；
- 3) 审计记录应包括事件的时间、类型、主客体标识和结果等，并保证审计记录不被删除或篡改；
- 4) 对记录数据进行分析，生成审计报表。

8.2 备份要求

城市生态环境数据备份要求包括但不限于：

- a) 除对数据本身进行备份外，还应备份数据配置信息、数据维护日志、系统访问日志及数据访问日志等；
- b) 定期对数据做增量备份及全量备份，数据备份应保存两个以上版本，全量备份应在访问量较小的时段进行；
- c) 支持手工备份和自动备份两种方式，备份策略配置灵活；
- d) 数据应能够在线备份，在不间断服务的情况下完成备份；
- e) 备份对象应能够按既定的备份策略备份到指定介质。

8.3 恢复要求

城市生态环境数据恢复要求包括但不限于：

- a) 备份数据应能方便快捷地恢复到在线系统，并确保其可用；
- b) 数据能够进行联机恢复，被恢复的数据应保持原数据的完整性和一致性；
- c) 重要数据应提供断点恢复功能，数据能够恢复到故障前的状态；
- d) 系统故障和数据丢失应在 4 h 内恢复正常运行；
- e) 对于数据库中数据块发生逻辑/物理损坏或单个表空间损坏的情况，能够支持通过恢复单个存储空间对数据库进行恢复；
- f) 数据库出现逻辑错误导致数据库无法正常使用时，确保能够采用数据库的全备份（或增量备份）结合数据库的归档日志文件进行恢复。

参 考 文 献

- [1] GB/T 19000—2016 质量管理体系 基础和术语
 - [2] GB/T 21337—2008 地理信息 质量原则
 - [3] GB/T 34960.5—2018 信息技术服务 治理 第5部分：数据治理规范
 - [4] HJ 1145—2020 生态保护红线监管技术规范 数据质量控制（试行）
 - [5] ISO 19157:2013 地理信息 数据质量（Geographic information—Data quality）
-