

DB3212

泰州市地方标准

DB3212/T 1147—2023

公共数据采集规范

Specifications for public data acquisition

2023-12-31 发布

2024-01-31 实施

泰州市市场监督管理局

发布

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。

本文件由泰州市大数据管理局提出、归口并组织实施、监督。

本文件由泰州市大数据管理局负责具体技术内容的解释。

本文件起草单位：泰州市大数据管理局、泰州市标准化院。

本文件主要起草人：陆建龙、陈书剑、李海鹏、孙慧、赵文涛、王小冬、陈书剑、刘小芳、吴薇、陈蓝生、王友成、张婧娴、郭健。

公共数据采集规范

1 范围

本文件规定了公共数据采集的基本原则、数据采集、数据安全保障等内容。

本文件适用于公共数据主管部门面向公共管理和服务机构，对公共数据进行收集、汇聚、整合。

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 38888 数据采集软件的性能及校准方法

3 术语和定义

下列术语和定义适用于本文件。

3.1

公共数据 public data

公共管理和服务机构为履行法定职责、提供公共服务收集、产生的，以电子或者其他方式对具有公共使用价值的信息的记录。税务、海关、金融监督管理等国家有关部门派驻泰州管理机构提供的数据。

3.2

公共数据平台 public data platform

以大数据、人工智能、区块链、物联网、云计算等技术为支撑，实现公共数据目录编制、归集治理、共享开放等全生命周期管理的技术平台。

3.3

公共管理和服务机构 public administration and service agencies

各级行政机关、法律法规授权的具有管理公共事务职能的组织、公共企事业单位。

3.4

数据采集 data acquisition

公共数据主管部门面向公共管理和服务机构，对公共数据进行收集、汇聚、整合的行为和过程。

4 基本原则

公共数据采集宜遵循以下基本原则：

——统筹性：统筹数据采集要求，建立省市区协同联动机制；

——合规性：按照法定职责和采集标准，获取公共管理和服务机构依法收集的数据；

——准确性：真实反映客观事物的状态；

——完整性：明确数据采集的范围、内容及目标，数据采集宜覆盖项目所规定的内容，不宜造成数据的缺失和遗漏；

——时效性：明确数据采集的频率和方式，实现数据的动态更新；

——安全性：加强公共数据采集过程中的安全保障。

5 数据采集

5.1 数据采集方案

- 5.1.1 采集方案应包括采集内容、采集频率、采集方式等。
- 5.1.2 公共管理和服务机构应当以数字化方式将收集、记录和存储的公共数据向市公共数据平台汇聚，非数字化数据在汇聚前应当进行数字化加工。
- 5.1.3 本行业、本领域公共数据可以由市级公共管理和服务机构统一向市公共数据平台汇聚。
- 5.1.4 公共数据平台应支持增量更新、全量更新、定时更新、事件触发更新和手动更新等方式。
- 5.1.5 公共数据平台提供多种标准协议的服务接入方式，包括但不限于数据库抽取、服务网关、消息队列、文件传输、直报系统、标准协议接口。
- 5.1.6 宜使用公共管理和服务机构的数据备份库作为采集对象，使用备份库时应保证数据一致性和可用性。对于海量数据，宜支持分批或增量读取，宜采用分布式方式对数据源进行读取。

5.2 在线采集

- 5.2.1 确定数据项间的逻辑关系，通过数据服务接口协议进行数据传输，实现对接采集。
- 5.2.2 数据协议方如发生数据变更，可根据业务需要申请增加数据信息项目，由公共数据平台依申请负责创建数据项。
- 5.2.3 公共数据平台应维护数据采集的业务管理规范，可以查询、查看采集数据信息，但应新增、删除数据信息。
- 5.2.4 公共数据平台方应确认写入数据与读取数据数量一致。判定一致后，按流程将数据入库待治理；发现数据存在差异，联系公共管理和服务机构重新获取加载信息，并将存储修改和原始记录，以供校验，获取差异数据应符合 GB/T 38888 的规定。

5.3 数据清洗

- 5.3.1 对原始数据开展数据清洗管理，定义清洗规则，分析、检查和处理数据源存在的质量问题。对定义的清洗方法的正确性和效率进行验证与评估，对不满足清洗要求的清洗方法进行调整和改进。
- 5.3.2 数据转换应对数据的标准代码、格式、类型等进行转换。必要时应建立数据转换规则表。

5.4 数据质量控制

对采集的原始数据进行清洗、转换、分析等处理，数据质量控制应符合以下要求：

- 完整性：应包含数据规则要求的数据的必要元素；
- 准确性：应真实反映数据所描述的实体；
- 一致性：应保证数据与其他特定上下文中使用的数据无矛盾；
- 时效性：应保证数据发生变化后及时被更新；
- 可访问性：应保证数据在需要时能被安全访问；
- 可追性：应保证数据能够被跟踪和管理。

6 数据安全保障

- 6.1 有关主管部门、公共数据平台方、公共管理和服务机构开展公共数据采集应建立健全全流程数据安全管理制度，落实公共数据安全保护责任，保障数据安全。
- 6.2 公共数据平台应建立公共数据资源安全管理制度，分级管理、按流程申请、按流程修改。确保数据来源的合法性、正当性、可定位溯源性，满足公共数据资源采集审计及监测的要求。
- 6.3 公共数据平台方应建立数据访问权限管理制度，明确公共数据平台方和数据协议方分级权限安全管理职责分工和工作要求，对数据采集环境、设施和技术采取安全管控措施，数据在整个采集、转化、传输过程中应依据授权使用，不被非法冒充、窃取、篡改、抵赖。
- 6.4 公共数据平台方应建立公共数据安全监测、安全审计、数字签名、数据加密等技术或手段，对不同数据进行分类并标识。采用安全技术进行安全维护。
- 6.5 公共数据平台方应建立应急事件响应机制，编制应急预案，开展应急演练，编制应急演练报告，持续优化应急预案等。
- 6.6 公共数据平台定期对公共数据采集的安全性进行风险评估，并据此制定相应的风险处理计划，及时排查安全漏洞、加固安全技术。

- 6.7 公共数据平台方开展公共数据管理安全培训，编制公共数据安全培训规划，培训公共数据专业知识，开展公共数据安全工作经验交流，宣传公共数据安全知识等。
- 6.8 公共数据平台应明确数据采集过程中个人信息和重要数据的知悉范围和安全管控措施。并采取必要的技术手段和管理措施保证数据不被泄露。
- 6.9 公共管理和服务机构应梳理接入应用系统和终端的公共数据，确定接入终端数量、网络带宽和接入地点 等信息，制定接入实施方案，明确访问控制、接入认证和审计等安全措施。

7 评价与改进

- 7.1 应建立公共数据采集的评价机制，对数据采集的效果进行综合分析评价，并不断改进。
- 7.2 评价可采用自我评价、用户满意度评价、第三方评价或多方评价相结合等方式进行。

参 考 文 献

- [1] 省政府办公厅关于印发江苏省政务“一朵云”建设总体方案的通知（苏政办发〔2023〕36号）
 - [2] 泰州市公共数据管理办法
-