

SZSD

数 字 山 东 工 程 标 准

SZSD07 0002—2024

数据要素技术与管理规范

Data elements technology and management Specification

2024 - 04 - 15 发布

2024 - 06 - 01 实施

潍坊市大数据局 发 布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 相关方责任 1

6 数据汇聚 2

 6.1 概述 2

 6.2 数据来源 2

 6.3 数据汇聚方式 2

 6.4 汇聚要求 2

7 数据治理 3

 7.1 概述 3

 7.2 治理要求 3

8 数据存储 3

 8.1 数据库分类 3

 8.2 存储要求 4

9 数据共享交换 4

10 数据分析与服务 4

 10.1 概述 4

 10.2 服务内容 4

11 数据资产 4

 11.1 概述 4

 11.2 元数据管理 4

 11.3 数据资产类目 5

 11.4 数据资产地图 5

 11.5 数据资产应用 5

12 数据安全 5

13 数据质量 5

 13.1 质量目标 5

 13.2 质量要求 5

参考文献 7

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。

本文件由潍坊市大数据局提出并归口。

本文件起草单位：潍坊市大数据局、潍坊市大数据中心。

本文件主要起草人：丁艳艳、王丽君、牟宁芳、柴力、潘洪华、徐兵、王春燕、李志臣、韩鹏东。

数据要素技术与管理规范

1 范围

本文件规定了数据要素的汇聚、治理、存储、共享交换、分析与服务要求，以及数据资产管理、数据安全和数据质量要求。

本文件适用于大数据工作主管部门、数据提供方和数据需求方的数据汇聚、治理、存储、共享交换、分析与服务等活动。

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 22239—2019 信息安全技术 网络安全等级保护基本要求
- GB/T 35273—2020 信息安全技术 个人信息安全规范
- GB/T 39477—2020 信息安全技术 政务信息共享 数据安全技术要求
- GB/T 41479—2022 信息安全技术 网络数据处理安全要求
- ISO/IEC 13249 信息技术 数据库语言 SQL多媒体和应用包

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据提供方 data provider

在数据汇聚、共享和开放过程中，提供数据（或接口服务）的公共管理和服务机构。

3.2

数据需求方 data user

在数据汇聚、共享和开放过程中，申请使用数据（或接口服务）的公共管理和服务机构。

3.3

资源目录 resource directory

依据规范的元数据描述，按照一定分类方法进行排序和编码的一组信息，用于描述信息资源特征，以便对信息资源的检索、定位与获取。

4 缩略语

下列缩略语适用于本文件。

- AI：人工智能（Artificial Intelligence）
- API：应用程序接口（Application Program Interface）
- BI：商业智能（Business Intelligence）
- HTTP：超文本传输协议（HyperText Transfer Protocol）
- JSON：基于JavaScript语法子集的开放标准数据交换格式（JavaScript Object Notation）
- REST：表述性状态传递（Representational State Transfer）
- SOAP：简单对象访问协议（Simple Object Access Protocol）
- SQL：结构化查询语言（Structured Query Language）
- WSDL：网络服务描述语言（Web Service Description Language）
- XML：可扩展标记语言（Extensible Markup Language）

5 相关方责任

数据要素管理涉及大数据工作主管部门、数据提供方和数据需求方，其相关职责包括：

- 大数据工作主管部门负责审核数据需求，完成数据需求方和数据提供方的对接；
- 数据提供方按照审核要求提供数据资源；
- 数据需求方按照审核要求使用数据资源，并确保数据安全。

6 数据汇聚

6.1 概述

数据汇聚是利用技术手段将分散于不同部门或不同系统的数据，实现统一获取和使用的过程。公共管理和服务机构应根据资源目录将本部门所有可汇聚的数据汇聚至潍坊市一体化大数据平台，并负责数据的更新；对于不予汇聚的数据，应向大数据工作主管部门提供书面说明，并征得大数据工作主管部门的同意。

6.2 数据来源

汇聚数据的来源包括但不限于：

- a) 政务数据：公共管理和服务机构履职和服务过程中采集、产生和管理的各类数据；
- b) 物联网数据：城市管理中各种物联网感知设备设施获取的数据，如环境感知、智能家居、楼宇感知信息、汽车与交通运输设备感知信息等；
- c) 视频数据：视频监控设备获取的视频及其相关数据，包括实时视音频、历史视音频、视频片段、图像等；
- d) 互联网数据：公共管理和服务机构利用技术手段，从网络、微信平台或其他行业大数据平台采集获得的互联网数据，包括互联网发布的发展趋势、白皮书，政府网站对外公开的信息报告、年度报告等。

6.3 数据汇聚方式

根据业务系统成熟度、技术能力以及接入成本等因素，选择合理的数据汇聚方式。数据汇聚方式包括但不限于：

- a) 库表接入；
- b) API 接口接入；
- c) 文件接入。

6.4 数据汇聚要求

6.4.1 基本要求

按照“一数之源、一源多用”的原则，实现数据资源一次采集、共享使用。除法律、法规另有规定外，应通过共享方式获取的政务数据，不得重复采集、多头采集。非涉密数据应汇聚至一体化大数据平台，并做好数据维护和更新。

公共管理和服务机构应根据资源目录将本部门所有可汇聚的数据汇聚至潍坊市一体化大数据平台，并负责数据的更新；对于不予汇聚的数据，应向大数据工作主管部门提供书面说明，并征得大数据工作主管部门的同意。

6.4.2 库表接入

数据提供方提供的库表接入方式应满足以下要求：

- a) 数据库的库名、表名、列名等应符合 ISO/IEC 13249 要求；
- b) 参与交换的数据库表之间不能有外键，表与表之间在物理上是独立关系，且每张表应有主键；
- c) 应优先配置国产数据库；
- d) 保存采集的日志记录。

6.4.3 API 接口接入

数据API接口接入方式宜满足以下要求：

- a) 支持 REST 和 Web Service 等接口类型；
- b) 采用 HTTP/S 1.0/1.1 服务传输协议；
- c) Web Service 服务消息封装协议采用 SOAP1.1/1.2，REST 服务消息封装协议采用 HTTP 1.0/1.1；
- d) 采用 XML、JSON、JAVA 对象等数据返回格式；
- e) 服务描述应包含服务调用者使用的具体访问接口内容，包括数据类型定义、操作、交互消息格式和内容、端口类型及绑定等，内容描述符合 WSDL1.1 的规定。

6.4.4 文件接入

文件接入方式应满足以下要求：

- a) 支持文件的读取、接收、校验、传输以及断点续传；
- b) 具备传输全过程监控能力，并保证传输的安全性、准确性和一致性；
- c) 对于异常文件的采集提供手工或自动采集手段。

7 数据治理

7.1 概述

对多源数据进行数据质量稽查、问题数据修复、数据标准化转换等处理，并基于目标数据模型进行数据关联和数据匹配融合，形成规范、一致的数据。

7.2 治理要求

数据治理要求如下：

- a) 大数据工作管理部门负责组织开展本地区的数据治理工作；
- b) 公共管理和服务机构应按照本级大数据工作管理部门的要求开展本部门数据治理，保证数据的准确性、完整性；
- c) 数据相关方（包括数据需求方）应在发现数据问题的 2 个工作日内，将问题反馈至数据提供方，数据提供方应在 5 个工作日内完成问题处理。

8 数据存储

8.1 数据库分类

数据库包括但不限于基础数据库、主题数据库、专题数据库。各类数据库宜包含以下数据。

- a) 基础数据库，是指将分散在各公共管理和服务机构的基础信息数据汇聚成库，为全社会提供全面、精准权威的基础数据支撑，应包括但不限于人口、法人单位、空间地理、电子证照、公共信用等子数据库。各子数据库应包括以下数据资源：
 - 1) 人口库应包括基本信息、出生信息、死亡信息、婚姻信息、社保信息、健康信息、公积金信息、房产以及证书等各类数据资源；
 - 2) 法人单位库应包括企业基本信息、行政执法信息、企业生产经营信息、以及税务、司法、信用等各类数据资源；
 - 3) 空间地理库应包括矢量数据、影像数据、高程模型数据、三维模型数据、地名地址数据等基础地理信息数据；
 - 4) 电子证照库应包括出生医学证明、营业执照、建筑施工许可证、卫生许可证、结婚证等通用类证照资源；
 - 5) 公用信用库应包括自然人、法人和其他组织的基本信息、守信信息、失信信息等。
- b) 主题数据库，是指围绕某一主题应用场景，将关联的各公共管理和服务机构数据汇聚形成的数据库，支撑跨部门、跨领域数据应用，包括但不限于医疗健康、交通出行、生态环境等数据库。

- c) 专题数据库，是指各公共管理和服务机构整合本行业各层级数据资源形成的数据库，由各行业特定时期的专项业务资源构成，如道路数据库包括道路名、长度、宽度、密度、运载能力、类型、结构、途经居民点、交通状况（车流量、车速限制）、道路位置等。

8.2 存储要求

数据存储的基本要求如下：

- a) 应支持数据上传、数据下载、资源目录查看、资源目录创建、资源目录删除、权限修改等操作；
- b) 应支持标准、开放的数据访问 API 对数据进行操作；
- c) 应提供数据自动备份或手动备份功能；
- d) 应支持数据更新、删除等数据管理功能。

9 数据共享交换

数据共享交换应满足以下要求：

- a) 根据实际业务需求，提供服务接口、库表交换、文件交换等多种方式，依托一体化大数据平台开展数据共享交换；
- b) 明确共享交换的数据资源与资源目录之间的映射关系，确保共享交换的数据资源都已编目。

10 数据分析与服务

10.1 概述

数据分析与服务应具备指标管理、BI分析等能力，并根据业务需求设计业务模型、数据模型、算法模型，支撑数据统计、预警预测、关系图谱、专题分析等数据分析服务。

10.2 服务内容

应提供以下数据分析与服务内容。

- a) 地理信息服务：基于空间地理基础数据库，对外提供标准统一的地理信息服务如空间定位、地名地址匹配、地图服务等，空间地理基础库的更新及维护需要广泛整合来源于不同部门的地理信息实时数据，如建筑模型、街景影像、地下管线等，需扩大地理信息数据采集更新渠道，以保持地理信息库的持续完善与实时更新。
- b) 人工智能决策分析：通过对部门已建的信息化系统、计划新建的信息化系统进行分析解构，并分析各部门智慧化应用对于人工智能技术支撑的需求，经需求汇总分析得出结论；人工智能决策分析包括但不限于人脸识别、车辆识别等视频分析算法模型，以及结构解析、评价描述、分析预测、智能推荐等综合性算法。
- c) 人工智能服务：基于各类 AI 算法，实现智慧交通违法抓拍、稽查布控以及平安城市、综治维稳等人工智能服务。
- d) 数据服务：应提供完善的数据服务，包含数据采集、清洗、管理、共享交换、分析展示等数据全生命周期运营的数据治理及创新应用体系。

11 数据资产管理

11.1 概述

数据资产管理应对数据资产全生命周期进行规划设计、过程控制和质量监督，具备数据标签管理、元数据管理、数据资产类目、数据资产地图、数据资产应用等方面的管理能力。

11.2 元数据管理

元数据管理包括但不限于以下功能：

- a) 实现多种方式的元数据采集；

- b) 提供元数据分析功能，包括数据全生命周期分析等，实现对数据资产的维护和使用；
- c) 提供元数据生命周期和版本管理，包括元数据变更统计，并提供根据版本号、日期过滤功能，元数据分析功能、查看版本详情及不同版本间对比功能。

11.3 数据资产类目

应根据元数据形成数据资产类目，其主要功能包括但不限于：

- a) 实现新增、删除、展示数据资产类目；
- b) 提供主题数据资源目、部门数据资源目录。

11.4 数据资产地图

提供数据资产盘点入口，应能对来自不同系统、不同部门的数据，查看数据表的业务信息、结构信息、影响分析、血缘分析、数据探索、数据预览等内容。

11.5 数据资产应用

11.5.1 提供数据资产的分析功能，展示数据资产的挂接率、满足率等应用分析数据。

11.5.2 提供数据资产应用成效入口，可进入数据资产应用成效查看全部应用情况。

12 数据安全

数据安全应符合以下要求：

- a) 在数据汇聚、数据传输、数据处理、数据存储、数据共享交换等过程中，均应具备用户身份管理和访问控制机制，支持对相关报表、数据访问权限的管理，兼顾多级分域数据访问安全；
- b) 数据传输应符合 GB/T 39477—2020 中 6.2.4.4 规定的安全传输要求；
- c) 建立数据传输通道时，应先进行双向认证，保证数据传输双方安全可靠；
- d) 应确保敏感数据在使用、传输过程中的鲁棒性、机密性、完整性和不可抵赖性；
- e) 数据备份应支持数据纠删码技术；
- f) 数据备份可采用多重密钥对数据存储进行加密保护；
- g) 数据处理应符合 GB/T 41479—2022 中第 5 章的要求；
- h) 数据共享交换应符合 GB/T 39477—2020 中第 6 章的要求；
- i) 数据删除应符合 GB/T 35273—2020 中 8.3 的要求；
- j) 应建立管理制度并采取技术措施，保证个人隐私不被泄露；
- k) 宜建立数据安全分级保护机制，以实现数据隐私保护、安全加密、调用异常监测等；
- l) 宜按照实际数据分级安全保护要求合理设置数据安全等级，并提供相应数据安全保护措施数据质量。

13 数据质量

13.1 质量目标

数据质量目标应包括规范性、完整性、准确性、一致性、时效性。

13.2 质量要求

13.2.1 数据规范性

数据资源应符合标准或业务规则，相关要求如下：

- a) 数据内容应符合相关国家标准、行业标准、地方标准的要求，暂无标准的也应按照行业习惯给出准确的数据属性；
- b) 数据内容应符合相应的业务规则；
- c) 数据内容应符合相关国家标准、行业标准、地方标准的安全规定。

13.2.2 数据完整性

数据资源应表述全面，无记录或字段缺失，相关要求如下：

- a) 数据元素宜被完整赋值；
- b) 数据属性字段宜被完整赋值。

13.2.3 数据准确性

数据资源应能反映真实的实体，相关要求如下：

- a) 数据内容应同预期数据内容相符合；
- b) 数据类型、数据长度、取值范围、数据精度应同预期数据相符合；
- c) 特定字段、记录、文件或数据集不存在重复；
- d) 不存在无效的数据。

13.2.4 数据一致性

数据资源间应保持一致，相关要求如下：

- a) 应只有一个部门负责某项数据的采集、录入、审核和更新，确保“一数一源”；
- b) 不同应用或用户使用的同一数据应是一致的；
- c) 关联数据应是一致的。

13.2.5 数据时效性

数据资源应符合业务实效性需求，相关要求如下：

- a) 一体化大数据平台数据应与各公共管理和服务机构的数据更新周期保持一致；
- b) 宜采用增量方式对汇聚数据进行更新。

参 考 文 献

- [1] 《政务信息资源目录编制指南（试行）》
 - [2] 《关于进一步健全政务数据共享协调机制加快推进数据有序共享的实施方案》（鲁政办发[2021]13号）
 - [3] 《山东省公共数据共享工作细则（试行）》（鲁数发[2021]9号）
 - [4] 《山东省电子政务和政务数据管理办法》（山东省人民政府令第329号）
 - [5] 《山东省公共数据开放办法》（山东省人民政府令第344号）
-