

数 字 山 东 工 程 标 准

SZSD17 0001—2024

农业大数据采集与分类规范

Specification for Agricultural Big Data Collection and Classification

2024 - 04 - 15 发布

2024 - 06 - 01 实施

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 2

5 采集及分类原则 2

6 采集及分类方法 2

7 检验规则 3

附录 A..... 4

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。

本文件由菏泽市大数据局提出并归口。

本文件起草单位：山东省大数据中心、菏泽市大数据中心、单县大数据中心、菏泽单州数字产业发展有限公司、山东华丰数字科技有限公司

本文件主要起草人：王开国、吴方超、高树锴、贾康、周建、吴雨桐、张冉冉、张西宁、胡凤霞、时天霄、曹子建、石秋发、王莹、张新宇、平善让等。

农业大数据采集与分类规范

1 范围

本文件规定了农业大数据的采集和分类原则以及方法。
本文件适用于单县各级农业大数据的采集和分类。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是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。

GB37/T 4473—2021 农业大数据分类与编码规范
GB34/T 3473—2019 农业大数据信息资源目录管理
GB/T 4754-2017 国民经济行业分类

3 术语和定义

下列术语和定义适用于本文件。

3.1

农业大数据 `agricultural big data`

融合了农业地域性、季节性、多样性、周期性等自身特征后产生的来源广泛、类型多样、结构复杂、具有潜在价值的数据集合。

3.2

农业大数据标志符 `agricultural big data identifier`

用于标识农业大数据的唯一一组字符。

3.3

采集设备 `collecting device`

对采集点的农业生产、加工、管理、服务等信息进行自动获取的装置。

3.4

编码 `coding`

给事物或概念赋予代码的过程。

3.5

代码 `code`

表示特定事物或概念的一个或一组字符。

注：这些字符可以是阿拉伯数字、拉丁字母或便于人和机器识别与处理的其他符号。

4 缩略语

ETL Extract Transform Load 数据仓库技术

5 采集及分类原则

5.1 采集原则

数据的采集应遵循以下原则：

- （1）可靠性：采集的数据来源清楚、真实可靠，能客观地反映采集对象的真实情况；
- （2）完整性：数据内容完整、无缺失，能反映采集对象的全面情况；
- （3）实时性：采集的数据实时同步，数据具备较高的时效性；
- （4）准确性：采集的数据准确无误，数据具有较大的应用价值。
- （5）规范性：采集的数据格式符合分类与编码规范，便于数据处理。

5.2 分类原则

数据的采集应遵循以下原则：

- （1）稳定性：根据分类的标准，按照农业大数据稳定的分类特性作为分类的基础和依据，确保分类结果稳定统一；
- （2）实用性：从实际角度触发，按照统一标准进行农业大数据分类，具有可适用性及便利性，简单便利；
- （3）兼容性：分类标准及原则符合国家标准和地区标准的要求；
- （4）可扩充性：在类目的设置或层次的划分上，留有适当的余地，以保证农业大数据增加时，不打乱已建立的分类体系，可根据实际情况进行类目扩充，扩充的类目应分别符合类目的设置规则。

6 采集及分类方法

6.1 采集方法

6.1.1 人工采集

信息采集员讲历史纸质文档及其他电子文档形式的数据，按照标准格式，人工手动录入系统中。

6.1.2 自动采集

（1）数据库同步对接

对于直接开放数据库接口的数据，使用ETL自动抽取方式进行数据库交换，安装ETL数据抽取工具接入部门业务数据库，按照数据格式需求、时间要求，对数据进行抽取。

（2）接口方式

对于不允许直连数据库的数据源或者不允许改变原有应用系统的数据，需提供数据接口，将数据接口配置到平台中，其他需求部门可以通过平台直接获取到本部门的数据，可以进行数据的批量下载、比对、查询等功能。

（3）网络爬取

利用互联网引擎技术实现有针对性、行业性、精准性的农业大数据的数据抓取，形成数据库文件。

（4）物联网

通过对接的项目、合作单位安装在园区、农场里的采集装置及传感装置，通过有线、无线等网络途径，同步上传的数据。

（5）GPS、北斗、遥感、气象等

利用GPS、北斗系统、遥感、气象等技术，获取的地理信息、农业信息以及气象信息等涉农数据同步上传。

(6) 其他方式

除上述方式外，其他符合的方式方法。

6.2 分类方法

农业大数据采用标志符的分类构成方式，由1位字母+8位数组成，分为四级代码。其中字母为采集方式代码，前两位数为一级分类代码，三-四位数为二级分类代码，五至八位数为三级分类代码，具体见附录A。

7 检验规则

7.1 采集检验规则

数据采集后，按照同一标准进行整理输出，根据时效性、地区性、实用性等基本原则进行数据检验，如因地区差异、政策差异采集到的与本地区不相符的农业数据，则以本地区农业政策导向及其他标准为准则，将无效数据剔除，保证数据的因地制宜性。

7.2 分类检验规则

分类按照现有国家标准、地区标准及行业标准进行检验，保证现有分类与现行的国家标准、地区标准及行业标准相吻合，存在冲突的，按照层级顺序，执行当前国家标准、地区标准及行业标准。

附 录 A

(规范性)

农业行业分类代码表

附录 A. 1 农业行业分类代码表

代码	一级分类	二级分类	三级分类
01	农作物		
0101		粮食类	
01010001			稻
01010002			小麦
01010003			玉米
01010004			谷物类
01010005			薯芋类
01010006			豆类
0102		油料类	
01020001			大豆
01020002			油菜
01020003			花生
01020004			芝麻
01020005			向日葵
01029999			不另分类的油料
0103		水果类	
01030001			瓜果类
01030002			浆果类
01030003			柑橘类
01030004			核果类
01030005			热带类
01030006			坚果类
01039999			不另分类的果品
0104		蔬菜类	
01040001			叶菜类
01040002			根菜类
01040003			花菜类
01040004			茎菜类
01040005			果菜类
01049999			其他蔬菜及其制品

0105		食用菌类	
01050001			伞菌类
01050002			耳类
01050003			多孔菌类
01050004			虫草类
01050005			块类
01059999			不另分类的食用菌
0106		药材类	
01060001			植物类
01060002			动物类
01060003			矿物类
01069999			其他类
0107		其他类	
01070001			纤维类
01070002			糖料类
01070003			饮料类
01070004			饲料类
01070005			烟草类
01079999			其他类
0108		观赏植物	
01080001			草本花卉
01080002			木本花卉
01080003			藤本花卉
01089999			其他观赏植物
0109		不另分类的种子、种苗及繁殖材料	
01090001			粮食种子、种苗及无性繁殖体
01090002			油料种子、种苗及无性繁殖体
01090003			水果种子、种苗及无性繁殖体
01090004			蔬菜种子、种苗及无性繁殖体
01090005			食用菌孢子、子实体
01090006			药材类根茎、种子、种苗

01090007			其他作物种子、种苗
01090008			观赏植物种子、种苗及无性繁殖体
01099999			其他种子、种苗及繁殖材料
0199		其他种植业产品	
