



中华人民共和国粮食行业标准

LS/T 1820—2018

粮食大数据资源池设计规范

Specification for grain big data resource pool

2018-01-08 发布

2018-03-01 实施

国家粮食局 发布

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 总体要求	2
5 层次结构	3
6 数据构成	3
7 数据描述	5
8 数据接口	7
9 数据模型与存储体系	7
10 数据安全	9
11 备份与恢复	9
12 可扩展性	10
13 资源池管理	10
14 运行环境	10
15 设计指标	10
附录 A (规范性附录) 元数据内容	11
附录 B (资料性附录) 元数据值域代码	22
参考文献	25

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准由国家粮食局提出。

本标准由全国粮油标准化技术委员会(SAC/TC 270)归口。

本标准起草单位：南京财经大学、国家信息中心。

本标准主要起草人：张璐、伍之昂、曹杰、申冬琴、雷涛、林曦、徐枫、宦茂盛。

引 言

随着粮食生产、流通、仓储各环节中信息技术的广泛应用,各级粮食主管单位、各大型粮油仓储企业建设和运营的粮食信息化系统中积累了大量高价值的粮食信息数据资源。然而,这些宝贵的粮食数据资源分散于各种应用系统中,彼此结构不同,无法融合共享,存在数据重复和冲突等现象,并且访问接口众多,数据获取困难,难以直接服务于各种业务。粮食大数据资源池旨在汇聚和集成各粮食信息系统中的数据,建立数据丰富、格式规范、存储安全、访问高效的数据资源平台,以支撑基于大数据的各种粮食信息化服务。本标准的制定将规范粮食大数据资源池的规划、设计、开发、部署及应用。

粮食大数据资源池设计规范

1 范围

本标准规定了粮食大数据资源池设计的总体要求、层次结构、数据构成、数据描述、数据接口、数据模型与存储体系、数据安全、备份与恢复、可扩展性、资源池管理、运行环境、设计指标等。

本标准适用于粮食大数据资源池的规划、设计、开发、应用、培训和运行维护,支持与粮食行业其他信息系统的互联互通。

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。

- GB/T 2260 中华人民共和国行政区划代码
- GB/T 2659 世界各国和地区名称代码
- GB/T 2887 计算机场地通用规范
- GB/T 4880(所有部分) 语种名称代码
- GB/T 7108 数据和交换格式 信息交换 日期和时间表示法
- 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 信息系统安全等级保护基本要求
- GB/T 22240 信息安全技术 信息系统安全等级保护定级指南
- GB/T 26882.1 粮油储藏 粮情测控系统 第1部分:通则
- GB/T 30170 地理信息 基于坐标的空间参照
- GB 50174 数据中心 设计规范
- LS/T 1708.2 粮食信息分类与编码 粮食加工 第2部分:技术经济指标分类与代码
- LS/T 1711 粮食信息分类与编码 财务会计分类与代码
- LS/T 1712 粮食信息分类与编码 粮食贸易业务统计分类与代码
- LS/T 1713—2015 库存粮食识别代码
- LS/T 1802 粮食仓储业务数据元
- LS/T 1806 粮食信息系统网络设计规范
- LS/T 1807 粮食信息安全技术规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

粮食大数据资源池 grain big data resource pool

汇聚和集成各种粮食相关数据,提供安全、可靠、高效的数据存储,并对外提供粮食数据服务的资源平台。

3.2

结构化数据 structural data

能够组织成行列结构,可识别的数据。通常是一条记录,或者是被正确标记过的数据中的某一个字段,并且能够被精确地定位到。

3.3

非结构化数据 unstructured data

不具备统一的结构,不方便用二维逻辑表来表现的数据,包括文本、图像、声音、网页等。

3.4

元数据 metadata

关于数据的数据。

3.5

元数据元素 metadata element

元数据最基本的信息单元,每一个元数据元素都用一个包含若干属性的集合来描述。

3.6

复合元素 compound data element

由若干元数据元素,或者元数据元素与其他复合元素,或者若干其他复合元素共同组成的元数据元素,通常用来表示较高层次的概念。

3.7

数据类型 data type

对数据的有效值域及对该值域中的值所允许的操作的规定,例如数值型、布尔型、日期类型、文本类型、复合类型等。

3.8

数据集 dataset

由相关数据组成的可标识集合,是元数据的描述对象。

3.9

数据集类型 dataset type

根据数据集的结构或资源特征,对数据集所作的类型划分。

3.10

实体 entity

按一定结构组织起来的数据的集合,其结构可用一组属性来刻画。

3.11

分布式文件系统 distributed file system

管理的物理存储资源分布于若干节点且节点间通过计算机网络相连的文件系统。

3.12

横向扩展 scale-out

通过增加新设备而非升级设备部件的方式提升系统处理能力的方法。

4 总体要求

粮食大数据资源池规划、设计、开发和部署应满足以下要求:

- a) 满足部门业务需求;
- b) 具有可操作性、可靠性和可用性;
- c) 具有可持续性和可扩展性;
- d) 将数据保密和安全作为高优先项;
- e) 履行验证和授权功能;
- f) 数据安全应与资源池建设同步考虑;

- g) 基于开放的行业标准和采用成熟的主流技术；
- h) 资源池管理责任机制由主管、建设和运维单位相应确定；
- i) 资源池各类技术与设备的选用应遵循经济性要求，满足当前及未来一定时间内需求的前提下，宜选择性价比最好的设备及技术。

5 层次结构

粮食大数据资源池应以云平台的方式建立并提供服务，平台分为两级：国家级资源池和省级资源池，对应的要求如下：

- a) 国家级资源池对接国家级粮食信息系统，为省级资源池的上级平台，接收来自省级资源池中的数据，汇集全国范围内的粮食数据资源。
- b) 省级资源池对接省级粮食信息系统，汇集本省粮食数据，为国家级资源池的下级平台，定期将数据同步至国家级资源池中。涉粮央企的粮食大数据资源池可比照省级资源池设置。
- c) 各级资源池提供其对应范围内的数据服务，在下级资源池中数据满足服务需求时，应首先使用下级资源池服务。

6 数据构成

6.1 通则

粮食大数据资源池应汇集粮食生产、收购、仓储、加工、销售等过程中的各种相关数据，与粮食有关的宏观经济数据、管理数据，以及互联网公众舆情等数据。资源池中应包含但不限于 6.2~6.10 所规定的的数据类别。

资源池中的数据既包含结构化数据，也包含非结构化数据，如文本、图片、音视频等，资源池应支持上述数据的存储与管理，并针对未来可能产生的新型粮食数据提供足够的扩展能力。

6.2 粮食生产数据

6.2.1 物料数据

种植粮食所使用的物料的使用记录，如种子种类、来源、数量、生产批次，农药编码、农药名称、生产商、供应商、数量，肥料编码、名称、生产商、供应商、数量等数据。

6.2.2 田间环境数据

田间种植环节的环境参数，如水（高度、水质）、空气（温度、湿度、CO₂ 浓度、光照）、土壤（湿度、pH）等数据。

6.2.3 产量统计数据

全国及各地区耕地面积、农作物播种面积、粮食播种面积、粮食总产量、粮食商品量、粮食平均亩产等信息。

6.3 粮食收购数据

粮食补贴数据、粮食收购主体构成、粮食收购数量、质量、收购进度、检验结果、收购凭证、国有企业收购量、政策性粮食收购量等执行情况信息，以及 LS/T 1713—2015 中 D.1 关联信息表所包含的收购信息数据。

6.4 粮食仓储数据

6.4.1 基本信息数据

LS/T 1802 中粮食仓储业务数据元分类表中的基本信息数据。

6.4.2 粮食出入库数据

LS/T 1713—2015 中 D.4 关联信息表所包含的库存信息数据。

6.4.3 仓储作业数据

降水时间、方法,熏蒸时间、药剂、浓度,通风时间,仓库温度、湿度、水分、氧气、二氧化碳、磷化氢、储粮害虫及螨类等信息,以及 LS/T 1802 中仓储作业信息数据元所包含的数据。

6.4.4 粮情测控数据

LS/T 1802 中粮情检测数据元所包含的数据,以及 GB/T 26882.1 中规定的粮情测控系统设备的运行记录。

6.4.5 仓储账目数据

各粮库粮食进出库、保管以及清仓查库过程中涉及的账目数据,包括:账本、统计报表、统计台账、原始凭证、粮库保管账、统计表、会计账、银行资金账等数据

6.4.6 仓储视频数据

粮仓视频监控系统产生的视频数据。

6.5 粮食加工数据

加工企业信息,加工生产线编码、名称、加工产品名称,产品规格,加工数量,加工时间,检测时间,检验结论,检验员,产品生产日期等数据,以及 LS/T 1708.2 所包含的加工指标的完成情况数据。

6.6 粮食销售数据

6.6.1 粮食市场数据

粮食销售总量、国有企业销售量、政策性粮食竞价交易数据、终端市场价格、主要粮食制品价格等数据。

6.6.2 粮食消费信息

粮食消费总量、人均消费量、工业用粮、饲料用粮、种子用粮等数据。

6.6.3 粮食贸易数据

LS/T 1712 中包含的粮食零售相关数据。

6.7 粮食运输数据

粮食的出发地、目的地,运输过程中的车辆轨迹,运输过程中粮食存放环境、时间,操作人员,通过 RFID 出入库系统检测的粮食货位号、运输工具编号、吨粮袋号、地磅计量信息等数据。

6.8 宏观粮食数据

6.8.1 宏观统计数据

产量大县、粮食质量指标、人口总数、地区产值、地区财政收入、城镇居民人均可支配收入、农民人均纯收入等基本信息。

6.8.2 粮食库存数据

国有及非国有粮食经营企业库存、粮食转化企业库存、农户存粮、城镇居民存粮等数据。

6.8.3 国际市场数据

LS/T 1712 中包含的粮食进出口数据、主要国家经济指标、粮食期货指数、现货价格、国际主要粮食制品价格、世界银行农业相关数据、国际粮农组织相关数据、美国农业部相关数据等。

6.8.4 宏观调控信息

粮食政策法规,粮食供需平衡预测,粮食综合生产能力评价,储备粮轮换计划编制信息等数据。

6.9 粮食管理数据

6.9.1 涉粮企业信息

全国粮食仓储企业、粮食加工企业等信息。

6.9.2 行政许可业务管理信息

粮食收购资格许可、储备粮承储资格许可、仓储单位备案、熏蒸作业备案、经纪人备案、质量追溯等数据。

6.9.3 粮食财务数据

LS/T 1711 中所包含的粮食财务会计数据。

6.10 其他外部数据

其他涉及粮食行业的外部互联网数据,包括各种与粮食相关的新闻报道、公众话题等。

7 数据描述

7.1 通则

粮食大数据资源池中的数据应通过元数据对其组织结构、管理信息、访问信息等内容进行描述。

元数据的描述对象为数据集,描述内容参见 7.2,应包含数据集描述信息、数据集分发与服务信息、结构描述信息 3 个必选模块,以及范围信息、联系信息 2 个可选辅助模块。

元数据元素的属性参见 7.3,应包含中文名称、英文名称、标识、定义、类型、值域、可选性、最大出现次数、注释等 9 个属性。

7.2 元数据内容

7.2.1 数据集描述信息

数据集描述信息模块为必选模块,可引用“范围信息”“联系信息”两个辅助模块中的内容。

数据集描述信息模块应记录数据集的基本信息。利用数据集描述信息模块中的元数据元素,能够

对数据集本身的基本属性信息进行详细描述,例如数据集名称、分类、创建者、摘要、来源、所采用的语言等,数据集用户能够通过浏览这些描述信息对数据集形成一个总体了解。

数据集描述信息包含的元数据元素见附录 A 的 A.1。

7.2.2 数据集分发与服务信息

数据集分发与服务信息为必选模块,可引用“联系信息”辅助模块中的内容。

数据集分发与服务信息模块应记录与数据集发行、获取及提供服务有关的信息,主要包括数据格式、使用数据集的技术要求、收费策略、权限声明、订购指南、联系信息等内容。数据集分发与服务信息模块还应记录与数据集相关的服务的标识信息,以及详细的服务元数据信息,该模块用于在数据集与相关服务之间建立连接。

数据集分发与服务信息包含的元数据元素见 A.2。

7.2.3 结构描述信息

结构描述信息为必选模块。

结构描述信息模块应记录数据集所包含实体的结构的有关信息,包括实体基本信息、属性信息、约束条件等内容。实体结构信息可包括关系型数据库中的数据表及其属性信息,以及数据字典中的信息。对于不通过关系型数据库管理,但具备关系数据库表格类似特征的实体,也可通过此模块来描述,例如具有明确字段或键的文档数据库和文本文件等。

结构描述信息包含的元数据元素见 A.3。

7.2.4 范围信息

范围信息为辅助模块,不得单独使用,而应被必选模块的有关元素引用。

范围信息模块应记录数据集内容所涉及的分、时间和空间范围的信息。

范围信息包含的元数据元素见 A.4。

7.2.5 联系信息

联系信息为辅助模块,不得单独使用,而应被必选模块的有关元素引用。

联系信息模块应记录与数据集有关的个人和组织的联系信息。

联系信息包含的元数据元素见 A.5。

7.3 元素属性

元数据元素(包括复合元素和数据元素)由包含 9 个属性的集合描述,集合包含的属性见表 1。

表 1 元素属性

属性名称	说明
中文名称	元素的中文名称
英文名称	元素的英文名称
标识	用字符串表示的元素标识
定义	对元素含义的解释
类型	元素所属的数据类型,见表 2
值域	元素值的允许范围,自由文本表示无限制
可选性	元素是必选元素还是可选元素,可能的值包括:M(必选)或 O(可选)

表 1 (续)

属性名称	说明
最大出现次数	元素所允许的出现次数,只出现一次用“1”表示。重复出现用“N”表示,不为1的固定出现次数可用相应的数字表示,如“2”“3”“4”,不出现用“0”或“NULL”表示
注释	对元素的补充说明、建议及其他

表 2 数据类型

数据类型	说明
文本	自由文本,表明对数据项(字段)的内容没有限制
数值	通过数值的形式表达值的类型,包括整数和实数等
时间	以年月日时分秒的形式表达的值的类型,参见 GB/T 7408
布尔	两个而且只有两个表明条件的值,如 True 或 False,1 或 0
二进制	通过二进制格式存储对象,如图片、音频、视频等
复合	由若干数据类型组成,或具有特定值的枚举类型

8 数据接口

8.1 数据获取接口

数据获取接口主要用于收集数据。在权限范围内,数据获取接口应支持通过如下方式收集数据:

- 粮食信息系统 API;
- 传感设备;
- 网络爬虫;
- 人工录入。

8.2 数据访问接口

数据访问接口用于对外提供数据访问服务。数据访问接口应支持如下访问方式:

- 数据库连接访问;
- 中间件访问;
 - Web service 标准接口;
 - Restful API;
 - FTP 文件下载;
- 第三方软件访问,采用此种方式时,应提供具体的接口访问协议说明和数据格式。

8.3 数据同步接口

数据同步接口在国家级和省级资源池间同步数据。省级资源池应定期通过同步接口向国家级资源池推送更新数据。数据同步接口应支持数据的增量更新,支持数据的压缩和加密传输。数据应封装为多个单元进行传输,每个传输单元的数据量不宜超过 5MB。

9 数据模型与存储体系

9.1 数据模型

粮食大数据资源池中的数据应采用合理的存储模型进行组织。资源池应至少支持关系模型、键值

模型、列存储模型、文档模型和图模型 5 种数据组织和存储模型,不同类型的数据宜采用的存储模型如下:

- a) 结构化数据宜采用关系模型组织;
- b) 具有明确键值对的非结构化数据宜采用键值模型组织;
- c) 面向分析型查询,需要高效压缩和 I/O 效率的数据宜采用列存储模型组织;
- d) 大段文本,或具有嵌套环结构的数据宜采用文档模型组织;
- e) 具有或能够抽象为明确的点边结构的数据宜采用图模型组织。

9.2 存储体系

9.2.1 通则

粮食大数据资源池应提供服务器本地存储与专用存储设备相结合的存储体系,存储介质应以磁盘为主,支持磁盘阵列,可辅以固态硬盘对高频访问数据进行加速。各种粮食数据应根据其访问频率和安全性要求存放于合适的存储位置:

- a) 服务器本地磁盘存储实时、高频访问的业务数据,宜采用分片方式使数据访问与存储能够在同一台服务器上进行,降低网络 I/O 开销。
- b) 外置专用存储设备存储备份数据,存储设备应支持对数据进行定时、增量式的备份,并支持将备份数据导入业务系统。

不同来源的粮食数据具有不同的格式和存储要求,粮食大数据资源池应建立混合存储体系,包含关系型数据库、分布式文件系统、内存数据库、缓存数据库、键-值数据库、图数据库、多媒体数据库、数据索引等。

9.2.2 关系型数据库

关系型数据库主要用于结构化数据的存储,应支持分布式部署,并能够利用固态硬盘进行重要数据的随机读写加速。

9.2.3 分布式文件系统

分布式文件系统用于存储大文件数据。分布式文件系统中的存储节点应通过以太网互联,能够屏蔽底层细节,提供统一、透明的文件存取和数据管理功能。分布式文件系统不得依赖于具体的操作系统,应支持数据分块和自动复制,不得产生单点失效,应支持层级结构(目录树),具有设置并管理文件属性与访问权限的能力,具备容错功能。

9.2.4 内存数据库

内存数据库用于存储需要被频繁查询的粮食数据。内存数据库应能够分布式、跨平台部署,并支持数据持久化。分布式部署的内存数据库应支持主从复制、增量复制以及数据同步。

9.2.5 缓存数据库

缓存数据库主要对一定时间内访问过的粮食数据进行缓存,以减轻对主数据库频繁查询的压力。缓存数据库一般应将数据存储在内存在中,并面向数据读取进行优化。缓存数据库应能够从不同的数据库中加载数据,但不得替代真正存放数据的各类数据库。

9.2.6 键-值数据库

键-值数据库主要用于存放不涉及过多数据关系和业务关系的非结构化文本数据。键-值数据库应提供直接访问键-值数据的函数,具备类似关系数据库的单表查询功能,应支持数据索引的建立和数据压缩。

9.2.7 图数据库

图数据库主要用于存放具有点、边结构的图数据。图数据库应支持图的水平分割和超大规模邻接矩阵查询,支持路径搜索,能够使用高级图形遍历语言及脚本,能够通过使用键值和关系进行索引。

9.2.8 多媒体数据库

多媒体数据库主要用于存放图像和音视频数据。在进行数据组织时,一般应将原始的图像和音视频文件存放于分布式文件系统,并在多媒体数据库中存储对应的文件路径。多媒体数据库宜采用关系模型或键-值模型的方式组织。

9.2.9 数据索引

数据索引通过索引服务支持对大数据资源池中的数据进行快速查询。数据索引既可利用各类数据库自带的索引功能,也可通过第三方索引软件提供。采用第三方索引软件时,软件应支持分布式部署,能够提供全文检索,拥有完备的查询语法,具备高可扩展性和容错能力。

10 数据安全

资源池中的数据安全应符合 GB 17859、GB/T 22239、GB/T 22240 和 LS/T 1807 中对数据安全的要求。其安全性要求还应包括:

- a) 数据库的安全应符合 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) 能够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;
- c) 能够对用户进行分组管理,并基于用户组进行授权;
- d) 能够利用安全协议与证书对用户进行认证;
- e) 能够通过日志记录用户对资源的访问行为,并提供审计功能;
- f) 能够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。

11 备份与恢复

11.1 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) 应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) 能够支持手工备份和自动备份两种方式,备份策略能够灵活配置;
- d) 数据应能够在线备份,在不中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,备份介质包含磁带、磁盘、磁盘阵列、光盘等;
- f) 在允许的情况下,可通过异地备份机制进一步提升不可抗力下的系统容灾能力。

11.2 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,被恢复的数据必须保持原数据的完整性和一致性,提供完整的系统数据安全监控、报警和故障处理;
- 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) 任何原因导致的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) 对于数据库出现逻辑错误导致数据库无法正常使用的情况,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。

12 可扩展性

资源池的扩展应以横向扩展为主。

13 资源池管理

应提供可视化管理界面,具备以下管理功能:

- 能够进行数据的导入和导出;
- 能够对资源池中设备的使用状态进行实时监控;
- 能够对服务器存储配额进行管理;
- 能够设置访问权限及各种数据安全和备份参数。

14 运行环境

资源池使用的服务器、专用存储设备及其他附属硬件设施应部署于专用机房中,所在机房场地应符合 GB/T 2887 的要求,机房设计应符合 GB 50174 的要求,设备间的组网应符合 LS/T 1806 的要求。

15 设计指标

在网络条件未造成瓶颈的前提下,大数据资源池应满足以下设计指标:

- a) 国家级资源池的容量不低于 1 PB,数据访问接口支持访问并发数不低于 10 000,在最大并发数下,用户的请求涉及数据库处理的响应时间最长不超过 3 s,文件传输速率不低于 20 MB/s。
- b) 省级资源池的容量不低于 50 TB,数据访问接口支持访问并发数不低于 3 000,在最大并发数下,用户的请求涉及数据库处理的响应时间最长不超过 3 s,文件传输速率不低于 10 MB/s。
- c) 系统消息处理延迟(从系统收到请求消息到返回应答的时间间隔)不大于 500 ms(99% 概率)。
- d) 省级资源池与国家级资源池间数据同步速率不低于 1 GB/s。
- e) 各级资源池应能够 7×24 小时不间断工作,省级资源池平均无故障率不低于 99%,国家级资源池平均无故障率不低于 99.9%。

附录 A
(规范性附录)
元数据内容

A.1 数据集描述信息

数据集描述信息包含的元数据元素见表 A.1。

表 A.1 数据集描述信息

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
1	数据集名称	Dataset Title	DatasetTitle	数据集的中文名称	复合		M	1	山 1.1.1~1.4 元素构成
1.1	数据集代码	Identifier	Identifier	标识数据集的唯一代码	文本	自由文本	M	1	
1.2	数据集中文名称	Chinses Title	CNTitle	数据集的中文名称	文本	自由文本	M	1	
1.3	数据集英文名称	English Title	ENTitle	数据集的中文名称	文本	自由文本	M	1	
1.4	数据集别名	Alias	Alias	数据集的其他名称	文本	自由文本	O	N	简称、俗称、缩写等
2	主题	Subject	Subject	描述数据集内容的词语或短语	复合		M	N	由 2.1~2.2 元素构成
2.1	主题词	Subject Words	SubjectWords	描述数据集的专业主题词语	文本	粮食学科主题词表中的专业词汇	M	N	
2.2	自由关键词	Keywords	Keywords	由用户自由选择的描述数据集内容的词语	文本	自由文本	O	N	
3	描述	Description	Description	对数据集内容的文本描述	文本	自由文本	M	1	摘要、目录、图形的文字说明等
4	目的	Purpose	Purpose	对开发该数据集的目的的说明	文本	自由文本	O	1	

表 A.1 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
5	类型	Type	Type	数据集中数据所属类型的说明	复合	表 B.1	M	1	
6	存储模型	Model	Model	数据集中数据所使用存储模型的解释	复合	表 B.2	M	1	
7	数据量	Size	Size	数据集所包含数据量的说明	复合		O	1	由 8.1~8.2 元素构成
7.1	记录数	Records Number	Records Num	数据集所包含的记录数	数值	非负整数	O	1	记录为最基本的数据粒度
7.2	存储量	Memory Size	Memory Size	以数据集所占物理存储空间表示的数据量	文本	自由文本	O	1	单位字节、KB、MB、GB、TB 等
8	数据来源	Source	Source	数据集中数据的来源	文本	自由文本	O	N	产生数据集中数据的粮食信息系统
9	数据集提供者	Provider	Provider	提供数据集的个人或组织名称	文本	自由文本	M	N	可调用数据集联系信息模块
10	数据集贡献者	Contributor	Contributor	对数据集创建做出贡献的个人或组织名称	文本	自由文本	M	N	可调用数据集联系信息模块
11	更新频率	Update Frequency	Frequency	描述数据集在多长的时间内更新一次	文本	表 B.3	O	1	
12	数据集时间	Dataset Date	Date	与数据集的创建和修改时间相关的时间	复合		M	1	由 12.1~12.2 元素构成
12.1	创建时间	Creation Time	Creation Time	数据集内容的创建日期	时间	GB/T 7408	M	1	

表 A.1 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
12.2	最近修改时间	Last Modification Time	LastModified	数据集内容最近一次修改时间	时间	GB/T 7108	O	1	
13	语种	Language	Language	数据集内容所采用的语种	文本	GB/T 4880	M	N	
14	关联数据集	Relation	Relation	与当前数据集相关的其他数据集	复合		O	N	由 16.1 ~ 16.3 元素构成
14.2	关联 URI	Related URI	RelatedURI	关联数据集的 URI	文本	自由文本	O	1	
14.3	关联类型	Relationship	Relationship	关联数据集与当前数据集之间的关系类型	文本	表 B.4	O	1	
15	数据集范围	Coverage	Coverage	数据集内容所涉及的范围和时空范围	复合		O	1	参见范围信息

A.2 数据集分发与服务信息

数据集分发与服务信息包含的元数据元素见表 A.2。

表 A.2 数据集分发与服务信息

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
1	权限声明	Right Statement	Right	对数据集的访问、使用和传播等行为有关的限制声明	文本	自由文本	O	1	

表 A.2 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
2	技术要求	Technology Requirement	Tech	使用该数据集必要的技术前提	文本	自由文本	O	1	
3	收费策略	Charge Policy	Charge	对数据集收费机制的描述	文本	自由文本	O	1	
4	订购指南	Order Direction	Order	有关数据订购方法或订购过程的说明	文本	自由文本	O	1	
5	访问时间	Access Time	Access	数据集提供访问服务的范围	文本	自由文本	O	1	
6	服务指示信息	Service Indiciary	SerIndiciary	经由服务元数据规范所描述的数据集为用户'提供服务的指示信息	复合		M	1	山 6.1 ~ 6.16 元素构成
6.1	服务唯一标识符	Service Identifier	SerIdentifier	服务的唯一标识符	文本	自由文本	M	1	
6.2	服务名称	Service Name	SerName	数据集所提供服务的名称	文本	自由文本	M	1	
6.3	服务类型	Service Type	SerType	数据集所提供服务的类型	文本	表 B.5	M	1	
6.4	连接服务 URI	Service URI	SerURI	访问服务的 URI	文本	自由文本	M	1	
6.5	数据库系统	Database System	DBSystem	存放数据集的数据库管理系统的名称	文本	自由文本	M	1	直接连接数据库时使用
6.6	数据库版本	Database Version	DBVersion	存放数据集的数据库管理系统的正式公开版本号	文本	自由文本	M	1	直接连接数据库时使用

表 A.2 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
6.7	数据库名称	Database Name	DBName	存放数据集的数据库的具体名称	文本	自由文本	M	1	直接连接数据库时使用
6.8	中间件系统	MiddleWare System	MidSystem	访问数据集时所使用的中间件系统的名称	文本	自由文本	M	1	利用中间件访问时使用
6.9	中间件版本	MiddleWare Version	MidVersion	访问数据集时所使用的中间件系统的正式公开版本号	文本	自由文本	M	1	利用中间件访问时使用
6.10	WWW 服务 URL	WWW URL	WWWURL	WWW 服务对外提供的网址	文本	自由文本	M	1	利用 Web 访问时使用,符合域名规则
6.11	FTP 匿名性	Anonymity	Anonymity	FTP 是否支持匿名服务	布尔	布尔值	M	1	利用 FTP 访问时使用
6.12	用户名	User Name	User Name	访问数据集所需的用户名	文本	自由文本	M	1	
6.13	密码	Password	Password	访问数据集所需的密码	文本	加密文本	M	1	
6.14	连接主机 IP	Host IP	IP	数据集所在系统提供网络服务的本机 IP 地址	文本	自由文本	M	1	符合 IP 地址规则
6.15	端口号	Access Port	Port	数据集所在系统为数据集访问服务提供的服务端口	数值	非负整数	M	1	
6.16	服务描述	Service Description	SerDesc	服务的描述信息	文本	自由文本	O	N	介绍有关该服务的情况
7	数据格式	Data Format	DataFormat	数据的格式说明	文本	自由文本	M	N	数据如何解析的说明
8	联系方式	Contact	Contact	数据集管理者(组织)的联系信息	复合		O	N	参见联系信息

6 A.3 结构描述信息

结构描述信息包含的元数据元素见表 A.3。

表 A.3 结构描述信息

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
1	检索点	Entry	Entry	数据库中作为检索起点的主表或视图	文本	自由文本	M	1	
2	实体	Entity	Entity	数据集存储实体的有关信息	复合		M	N	
2.1	实体名称	Entity Name	EntityName	实体的存储名称	文本	自由文本	M	1	
2.2	实体中文标签	Entity Label	EntityLabel	实体文件的描述性中文标题	文本	自由文本	O	1	
2.3	实体类型	Entity Type	EntityType	实体的类型	文本	自由文本	O	1	
2.4	实体定义	Entity Definition	EntityDef	对实体所表示意义的解释和说明	文本	自由文本	M	1	
2.5	主键	Primary Key	PrimaryKey	对实体主键的声明	文本	自由文本	O	1	
2.6	属性	Attribute	Attribute	实体中属性的基本信息	复合		O	N	由 2.6.1~2.6.12 元素构成
2.6.1	属性名称	Attribute Name	AttributeName	属性在实体中存储时采用的名称	文本	自由文本	M	1	若实体类型为数据表,则此元素指的是表中字段名称
2.6.2	属性中文标签	Attribute Label	AttriLabel	属性的描述性中文名称	文本	自由文本	O	1	
2.6.3	属性类型	Attribute Type	AttriType	属性在实体中存储的数据类型	文本	自由文本	M	1	

表 A.3 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
2.6.4	存储长度	Length	Length	属性在实体中设定的最大存储长度	数值	非负整数	M	1	单位为字节
2.6.5	计量单位	Unit	Unit	计量属性值的基本单位	文本	自由文本	O	1	建议采用标准的计量单位名称或标识
2.6.6	精度	Precision	Precision	属性值的精确程度	数值	非负实数	O	1	以计量单位元素为单位
2.6.7	默认值	Default	Default	属性值的默认值	文本	自由文本	O	1	
2.6.8	属性描述	Attribute Descriptio	AttriDesc	对属性含义的描述性阐释	文本	自由文本	O	1	
2.6.9	关联实体	Related Entity	RelatedEntity	以外键为主键的相关实体	文本	自由文本	O	1	应填写相关实体的存储标识
2.6.10	关联类型	Relation Type	RelationType	实体之间关联关系所归属的种类	文本	自由文本	O	1	2.6.9 选用时,该元素必选,可一对一、一对多、多对多等关系
2.6.11	是否只读	ReadOnly	ReadOnly	属性是否为只读属性	布尔	布尔值	O	1	
2.6.12	可否空值	NullKey	NullKey	属性是否可以空值	布尔	布尔值	O	1	应填写非空的属性的标识
3	关系	Relationship	Relationship	数据集所存储实体之间的关系的基本信息	复合		O	N	由 3.1~3.5 元素构成
3.1	关联实体	Relation Entity	RelationEntity	实体关联关系中实体对象	文本	自由文本	M	1	
3.2	关联实体属性	Relation Entity Attribute	Relation EntityAttri	实体关联关系中,关联实体用来实现与子实体关联的属性	文本	自由文本	M	1	

表 A.3 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
3.3	关联子实体	Child Entity	ChildEntity	实体关联关系中实体对象	文本	自由文本	M	1	
3.4	关联子实体属性	Child Entity Attribute	Child EntityAttri	实体关联关系中子实体实现与关联实体关联的属性	文本	自由文本	M	1	
3.5	关联类型	Relation Type	RelationType	实体之间关联关系所归属的种类	文本	自由文本	M	1	可一对一、一对多两种关系

A.4 范围信息

范围信息包含的元数据元素见表 A.4。

表 A.4 范围信息

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
1	数据集内容范围	Taxonomy	Taxonomy	数据集内容所涉及的范围	复合	表 B.6	O	N	
2	数据集时间范围	Temporal Range	Temporal	数据集内容所涵盖的时间范围	复合		O	1	由 2.1~2.2 元素构成
2.1	起始时间	Begin Time	Begin	数据集原始数据生成或采集的起始时间	时间	GB/T 7108	M	1	
2.2	终止时间	End Time	End	数据集原始数据生成或采集的终止时间	时间	GB/T 7108	M	1	
3	数据集地理范围	Geographic Extent	GeoExt	数据集内容所涉及的地理区域	复合		O	1	由 3.1~3.4 元素构成

表 A.4 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
3.1	描述	Geographic Description	GeoDesc	有关地理区域的描述	文本	自由文本	O	1	
3.2	行政区划范围	Administrative Region Extent	ReExt	数据集内容所涉及的行政区划范围	复合		O	1	由 3.2.1~3.2.3 元素构成
3.2.1	国家	Country	Country	数据集覆盖的国家	文本	GB/T 2659	O	1	
3.2.2	省(市、自治区)	Province	Province	数据集覆盖的省、市、自治区	文本	GB/T 2260	M	1	
3.2.3	城市	City	City	数据集覆盖的城市	文本	自由文本	M	1	
3.3	经纬度范围	Longitude and Latitude Extent	LLExt	数据集覆盖的经纬度范围	复合		O	1	由 3.3.1~3.3.4 元素构成
3.3.1	最东经度	East Bound Longitude	EastBL	数据集覆盖范围最东边坐标,用十进制(东半球为正)	文本	自由文本	M	1	
3.3.2	最西经度	West Bound Longitude	WestBL	数据集覆盖范围最东边坐标,用十进制(东半球为正)	文本	自由文本	M	1	
3.3.3	最南纬度	South Bound Latitude	SouthBL	数据集覆盖范围最东边坐标,用十进制(东半球为正)	文本	自由文本	M	1	
3.3.4	最北纬度	North Bound Latitude	NorthBL	数据集覆盖范围最东边坐标,用十进制(东半球为正)	文本	自由文本	M	1	
3.4	垂直范围	Vertical Extent	VerExt	数据集覆盖的垂直范围	复合		O	1	由 3.4.1~3.4.3 元素构成
3.4.1	垂直最高值	Vertical Maximum Value	MaxVal	数据集包含的垂直范围最高值	文本	自由文本	M	1	

表 A.4 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
3.4.2	垂向最低值	Vertical Minimum Value	MinVal	数据集包含的垂向范围最低值	文本	自由文本	M	1	
3.1.3	垂向基准名称	Vertical Datum Name	Datum	垂向最大值和最小值的原点信息	文本	GB/T 30170	M	1	
1	组织机构范围	Organization Extent	OrgExt	数据集覆盖的机关、企事业单位及其他组织范围	文本	自由文本	O	1	

A.5 联系信息

联系信息包含的元数据元素见表 A.5。

表 A.5 联系信息

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
1	联系人名称	Contact Name	ConName	与数据集有关的联系人姓名或组织名称	复合		M	1	由 1.1~1.3 元素构成
1.1	个人名称	Individual Name	Individual	个人联系人姓名	文本	自由文本	O	N	
1.2	职务	Position	Position	个人联系人所担任的职务	文本	自由文本	O	N	
1.3	组织名称	Organization Name	OrgName	联系组织名称	文本	自由文本	O	N	
2	联系地址	Contact Address	ConAdd	联系人或组织的通信地址信息	复合				由 2.1~2.8 元素构成
2.1	国家	Country	Country	联系人或组织所在的国家名称	文本	GB/T 2659	O	1	

表 A.5 (续)

序号	中文名称	英文名称	标识	定义	类型	值域	可选性	最大出现次数	注释
2.2	省(市、自治区)	Province	Province	联系人或组织所在的省、市、自治区的名称	文本	GB/T 2260	M	1	
2.3	城市	City	City	联系人或组织所在的城市名称	文本	自由文本	M	1	
2.4	地址	Address	Address	联系人或组织所在的具体地址,从城市之后写起	文本	自由文本	M	N	
2.5	邮政编码	Postal Code	PosCode	联系人或组织所在地址的邮政编码	文本	自由文本	M	N	
2.6	传真	Fax	Fax	联系人或组织的传真号码	文本	自由文本	O	N	
2.7	电话	Phone	Phone	联系人或组织的电话号码	文本	自由文本	M	N	
2.8	电子邮件	Email	Email	联系人或组织所的电子邮件地址	文本	自由文本	M	N	

附 录 B
(资料性附录)
元数据值域代码

B.1 粮食数据集类型

粮食数据集类型见表 B.1。

表 B.1 粮食数据集类型

名称	代码	定义
结构化数据	001	具有明确字段结构的数据
键值数据	002	以键值对表示的数据
文本数据	003	以文本像是表现的数据
图数据	004	具有明确点、边结构的数据
轨迹数据	005	移动轨迹或静止点数据
二进制文件	006	以二进制形式表现的数据
图像	007	以图片或影像表现的数据
音频	008	以音频表现的数据
视频	009	以视频表现的数据
其他	010	以上类型不涵盖的数据

B.2 粮食数据存储模型

粮食数据存储模型见表 B.2。

表 B.2 数据存储模型

名称	代码	定义
关系模型	001	用二维表的形式表示实体和实体间联系的数据模型
键值模型	002	以键值对形式存储实体及其值的数据模型
列存储模型	003	将所有记录中相同字段的数据聚合存储的数据模型
文档模型	004	以文档及其嵌套结构存储实体的数据模型,文档中通常以键值对组织
图模型	005	以点和边描述和存储实体及其相互关系的数据模型

B.3 数据更新频率

数据更新频率见表 B.3。

表 B.3 数据更新频率

名称	代码	定义
不定期	001	不定期更新数据
每日	002	每日更新数据
每十天	003	每十天(旬)更新数据
每周	004	每周更新数据
半月	005	每半月更新数据
每月	006	每月更新数据
两月	007	每两月更新数据
每季度	008	每季度更新数据
每半年	009	每半年更新数据
每年	010	每年更新数据
两年	011	每两年更新数据
三年	012	每三年更新数据
按需要	013	定期但无法用上述名称描述的更新频率

B.4 关联类型

关联类型见表 B.4。

表 B.4 关联类型及其定义

名称	标签	代码	定义
被替代	Is Replaced By	001	所描述的数据集已被参照数据集所代替、替换或取代
替代	Replaces	002	所描述的数据集代替、替换或取代了被参照数据集
部分于	Is Part Of	003	所描述的数据集是被参照数据集物理上或逻辑上的一个组成部分
部分为	Has Part	004	所描述的数据集或者在物理上或者在逻辑上包括了被参照的数据集
被参照	Is Referenced By	005	被参照的数据集参考、引用或以另外的方式指引所描述的数据集
参照	References	006	所描述的数据集参考、引用或以其他方式指示了被参照数据集
格式转换于	Is Format Of	007	所描述的数据集与被参照数据集有相同的知识内容,但用了另一种格式表现出来
格式转换为	Has Format	008	所描述的数据集在被参照数据集之前出现,参照数据集在实质上与所描述的数据集有着相同的知识内容,只是格式不同

B.5 数据集服务类型

数据集服务类型见表 B.5。

表 B.5 数据集服务类型

名称	代码	定义
数据库连接	001	直接连接数据库访问数据集中的数据
中间件访问	002	通过中间件访问数据集中的数据
Web 服务	003	通过 Web 服务访问数据
FTP 服务	001	利用 FTP 获取数据文件

B.6 数据集分类

数据集分类见表 B.6。

表 B.6 数据集分类

名称	代码	定义
粮食生产数据	001	粮食生产过程中产生的种植、物料、环境及相关统计数据
粮食收购数据	002	粮食收购过程中相关数量、质量等数据
粮食仓储数据	003	粮食仓储中的仓库、环境、粮情、测控系统运行、视频监控等数据
粮食加工数据	004	粮食加工过程中的生产线、产品质量等数据
粮食销售数据	005	粮食销售中的市场和消费数据
粮食运输数据	006	粮食运输过程中的环境、车辆轨迹等数据
宏观粮食数据	007	宏观的粮食储量、经济、流通、调控等数据
粮食管理数据	008	涉粮单位信息、行政许可等数据
其他外部数据	009	与粮食相关的互联网舆情、新闻及其他数据

参 考 文 献

- [1] GB/T 21061—2007 国家电子政务网络技术和运行管理规范
 - [2] GB/T 30850.3—2014 电子政务标准化指南 第3部分:网络建设
 - [3] YD/T 1190—2002 基于网络的虚拟IP专用网(IP-VPN)框架
 - [4] 中国科学院数据库核心元数据标准 2.0
-