

# 公共数据管理 数据共享规范

## 1 范围

本文件规定了公共数据共享相关术语和定义、缩略语以及数据归集、数据共享、共享专题库/数据表和质量评估的要求。

本文件适用于公共数据在归集、共享、质量评估等环节的管理。

## 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 7408 数据元和交换格式 信息交换 日期和时间表示法

GB 11643 公民身份号码

GB 32100 法人和其他组织统一社会信用代码编码规则

DB33/T 2350-2021 数字化改革术语定义

DB33/T 2351 数字化改革 公共数据分类分级指南

## 3 术语和定义

DB33/T 2350-2021界定的以及下列术语和定义适用于本文件。

### 3.1

#### 公共数据 public data

国家机关、法律法规规章授权的具有管理公共事务职能的组织，在依法履行职责和提供公共服务过程中，所获取和产生的数据资源以及法律、法规规定纳入公共数据管理范围的其他数据资源。

[来源：DB33/T 2350-2021，定义3.2.2.1，有修改]

### 3.2

#### 数据高铁 high-speed data

通过数据库的日志采集、解析、入库等手段，提高公共数据传输速度、准确度、完整度的数据归集方式。

[来源：DB33/T 2350-2021，定义3.2.2.7]

### 3.3

**实时归集** real-time collection

通过数据高铁模式由日志同步实现数据实时归集的方式。

### 3.4

**数据仓** data warehouse

国家机关以及具有公共事务管理职能的组织采集和获取的，随时间动态更新，信息类型、格式等相对稳定的各类公共数据资源集合。

[来源：DB33/T 2350-2021，定义3.2.2.8]

### 3.5

**接口共享 interface Sharing**

通过接口进行数据共享的方式。

### 3.6

**大数据处理分析系统 max compute**

一种快速、完全托管的GB/TB/PB级数据仓库解决方案。提供了完善的数据导入方案以及多种经典的分布式计算模型，能够更快速的解决海量数据计算问题，有效降低成本，并保障数据安全。

### 3.7

**数源部门 data-source department**

共享数据提供部门。

### 3.8

**数据使用部门 data use department**

申请使用共享数据的部门。

### 3.9

**专题库/数据表 thematic data base/table**

对治理后的数据进行专项分类形成可应用的库/表。

### 3.10

**数据元 data element ; DE**

用一组属性规定其定义、标识、表示和允许值的数据单元。

[来源：GB/T 18391.1-2009，定义3.3.8]

## 4 缩略语

下列缩略语适用于本文件。

**HTTPS:** 以安全为目标的HTTP通道，在HTTP的基础上通过传输加密和身份认证保证了传输过程安全性 (Hyper Text Transfer Protocol over Secure Socket Layer)

**SSL:** 安全套接字协议 (Secure Sockets Layer)

**IRS:** 一体化资源系统 (Integrated Resources System)

RDS: 云数据库 (Relational Database Service)

## 5 数据归集

### 5.1 数据资源体系架构

数据资源体系架构见图1。

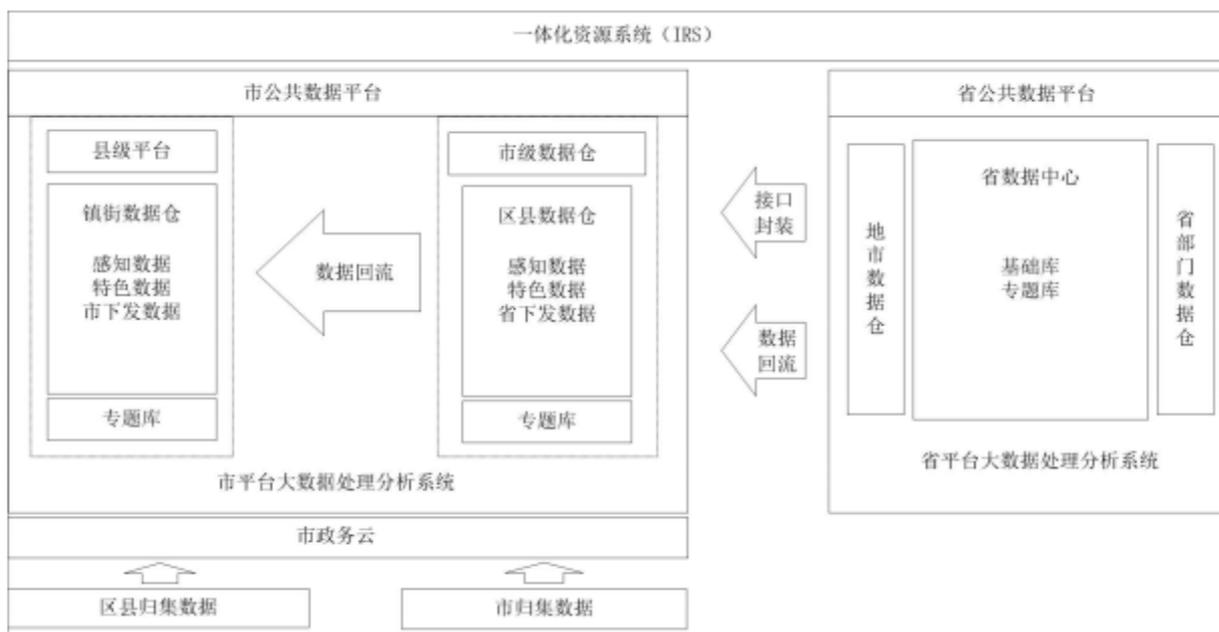


图 1 数据资源体系架构

## 5.2 归集要求

5.2.1 市级平台按照省市平台建设责任分工，建立完善的数据归集系统。

5.2.2 县级平台按需建设数据归集系统，建立本地数据归集能力，实现县域数据归集。

5.2.3 市县平台数据归集系统采用以下两种技术模式：

——数据交换模式：通过库表桥接方式实现数据归集，

——数据高铁模式：通过数据库的日志采集、解析、入库等手段，提高公共数据传输速度、准确度、完整度的数据归集模式。

5.2.4 数据交换模式的交换体系应包含数据交换系统、交换前置节点、交换中心、桥接、数据传输、交换监控等模块，应支持数据库、文件的归集方式。

5.2.5 市级平台的数据高铁系统应具有县级数据高铁专属通道和管理模块，以满足县级平台本地数据实时归集需求。

5.2.6 数据高铁模式的实时归集系统应包含数据采集系统工具部署、数据实时仓管理、运维管理系统。实时归集的实施包含对各业务系统的数据采集技术支持和系统对接联调。实时归集运维服务包含运维流程规范制定、监控搭建、技术整改、文档收集、运维评估等。

## 6 数据共享

### 6.1 共享方式

## 6.1.1 接口共享

6.1.1.1 接口共享系统实现市县两级复用，市接口共享系统与省接口共享系统实现对接和封装。市级平台应通过身份认证、权限管理、密钥认证、数据加密、白名单机制、HTTPS 协议、SSL 证书等技术措施，保障数据接口共享系统在数据使用与传输过程中的网络安全与数据安全。

统调用响应时间应低于 1500ms。高频调用接口可支撑每秒 500 次并发量；低频调用接口可支撑每秒 100 次并发量。

### 6.1.2 批量共享

#### 6.1.2.1 库表授权

市级公共数据平台提供统一的大数据分析处理系统环境，各应用部门通过库表授权方式，实现批量数据共享使用。处理后的结果数据根据数据内容的分类分级标准，经过对应审批后，通过生成接口或批量导出方式使用。

#### 6.1.2.2 数据回流

省、市、县级平台之间的批量数据回流：

- a) 省通过库表授权的方式，将省里涉及地市的数据授权给市级数据仓；
- b) 市通过库表授权的方式，将市里涉及县级的数据授权给县级数据仓。

## 6.2 数据共享体系

6.2.1 数据共享体系由数源部门、公共数据主管部门、数据使用部门组成。

6.2.2 公共数据主管部门负责平台建设、管理，数据编目、分类分级、归集、共享以及特色数据仓建设实施，负责本级数据使用部门共享数据申请的审批管理工作。

## 6.3 共享要求

### 6.3.1 共享申请

数据使用部门通过省 IRS 系统在线提出数据共享需求，要素应包括但不限于：

- 数据需求部门，
- 联系人，
- 联系方式，
- 应用场景，
- 申请理由，
- 需求数据项，
- 数源单位。

### 6.3.2 数据使用

市公共数据平台内的数据使用（见图2）可通过批量数据导出或接口服务方式将大数据分析处理系统内处理分析后的结果数据透出至云服务器或数据库：

- 批量导出：应提交数据导出申请至数据主管部门，数据主管部门审批通过后，由实施人员进行数据导出授权，实施完成后方可导出数据。导出后的数据不能用于申请范围以外的使用途径；

——接口导出：应提交接口申请至数据主管部门，数据主管部门审批通过后，由实施人员生成发布接口，生成接口过程中涉及的数据导出至 RDS 流程无需审批，实施完成后提供接口调用信息。

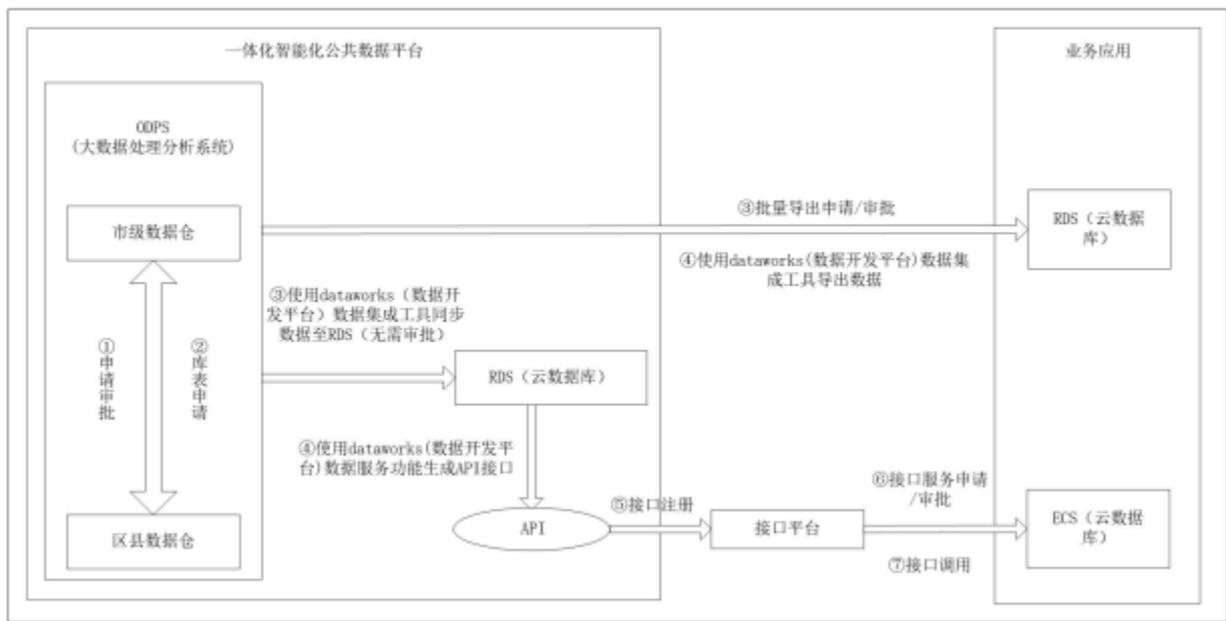


图 2 数据使用示意图

## 7 共享专题库/数据表建设

### 7.1 共享专题库的构成

#### 7.1.1 字典表

字典表是对各数据表字段要素信息的总体描述，应包括但不限于：

- 要素名称，
- 代码，
- 关联关系，
- 属性内容，
- 表示方法。

#### 7.1.2 代码表

字典类指标的输入值有特定的数据集合，即值域字典，一般包含由代码和含义组成的键值对，键值对的代码可定长也可不定长；代码由相关业务部门制定编码规则；字典类指标包含行政区划、性别、行业代码等。字典类指标应符合：

- 字典类字段应提供对应字典表，
- 字段存储值不应超出字典范围，
- 字段类型：VARCHAR。

### 7.1.3 业务数据表

应由核心业务数据元和扩展业务数据元组成，且能反映核心业务指标。

## 7.2 共享专题库/数据表表名命名规则

7.2.1 业务数据表表名命名规则：区域简称+部门简称+业务系统名+核心业务+表\_扩展。具体要求如下：

- a) 区域简称：规范化简称，
- b) 部门简称：规范化简称，
- c) 业务系统名：业务系统登记名，
- d) 核心业务：一般 4~6 个中文字，且能反映业务核心内容，
- e) 表\_扩展：表类型的说明，如字典表、代码表。

示例：宁波市交通局公交运营平台司机登记表、宁波市交通局公交运营平台司机登记表\_字典。

7.2.2 共享专题库命名规则：区域简称+核心业务+专题库\_扩展。具体要求如下：

- a) 区域简称：规范化简称，
- b) 核心业务：能反映一类相关核心业务的名称，
- c) 其他：英文的共享专题库/数据表表名在中文名称的基础上，采用首位拼音字母组合而成，如出现重名情况时，将核心业务部分采用中文名的全部拼音字母代替，如仍有重名，则通过增加扩展部分进行区分。

示例：宁波市司机专题库、宁波市营运车辆专题库。

## 7.3 业务表数据元组成规则

### 7.3.1 组成规则

7.3.1.1 业务表内的数据元由关键索引数据元、核心业务数据元、扩展业务数据元三部分组成。

7.3.1.2 关键索引数据元应符合附录 A 的要求；核心业务数据元应符合附录 B 的要求。

### 7.3.2 关键索引数据元

关键索引数据元的数据内容一般采用编码形式：

- 身份证件号码应符合 GB 11643 的要求，
- 统一社会信用代码应符合 GB 32100 的要求，
- 其他关键索引数据元编码应符合数源部门的编码规则。

### 7.3.3 核心业务数据元

核心业务数据元应描述业务的关键和基本属性。

### 7.3.4 扩展业务数据元

在核心业务数据元基础上，随着业务数据表的共享使用，所产生的扩展业务属性或关联业务索引数据元。

## 8 质量评估

### 8.1 评估维度

数据质量评估维度可从7个维度来评估数据是否达到预期设定的质量要求，主要内容如下：

- 完整性：用于度量数据丢失率或数据不可用率，
- 准确性：用于度量数据和信息的正确率，
- 及时性：用于度量数据更新的及时率
- 规范性：用于度量数据按统一格式、规范存储的比例，
- 唯一性：用于度量数据的重复率，
- 一致性：用于度量数据的值在信息含义上的不一致率，

——关联性：用于度量关联数据的缺失率。

## 8.2 评估方法

采用计算机方式进行评估，按照一定的编码规则对评估规则进行编码，评估规则编码定长4位，分为两段，第一段占1码位，为第1位，用大写字母A~F表示，第二段占3码位，为第2~4位，从001开始顺序编码，具体评估规则及编码见表1：

表 1 数据质量评估规则及编码

维度	编码	评估规则
A 完整性	A001	校验主要字段是否为空
B 准确性	B001	检验数值是否异常
	B002	校验数据含有TAB/ENTER键
	B003	检验字段首部是否含有空格
	B004	检验字段尾部是否含有空格
C 及时性	C001	检验字段数据更新的及时性
D 规范性	D001	校验字段长度
	D002	检验字段是否超出字典范围
	D003	检验字段是否含有特殊字符
	D004	检验字段是否含有给定字符以外的特殊字符
	D005	检验字段是否含有特定字符
	D006	检验字段是否含有除特定字符外的字符
	D007	检验字段格式是否规范
	D008	检验字段是否符合编码规则
E 唯一性	E001	校验主键是否重复
	E002	校验记录是否重复
F 一致性	F001	检验字段A与字段B是否一致
	F002	检验字段记录的统一性
G 关联性	G001	检验字段A与字段B逻辑关系

## 8.3 计算方法

通过公式1计算得到每条规则的分值，利用评估模型得到数据集的数据质量分：

$$F_j = \frac{\sum_{i=1}^n W_i * S_i}{\sum_{i=1}^n W_i} \dots\dots\dots (1)$$

式中：

$F_j$  ——数据 $D_j$  的质量分；

$W_j$  ——规则 $R_i$  的权重；

$S_i$  ——规则 $R_i$  的分值。

## 附 录 A

（规范性）

### 公共数据信息类别及关键索引数据元

公共数据元经按业务内容分类，由两段3位编码进行标识。第一段三位数字为一级分类，第二段三位数字为二级分类。各类别信息的关键索引数据元见表A.1。

示例：“自然人职业信息”的类别标识符为“001002”，其中前三位为“001”为所属一级分类，即“自然人基本信息类”，后三位“002”为在二级分类中的顺序号。

**表 A.1 信息类别及关键索引数据元表**

标识符	信息类别		关键索引数据元
001001	自然人基本信息类	自然人基本信息	身份证件号码
001002		自然人职业信息	工作单位统一社会信用代码、身份证件号码
002001	机构团体类	机构基本信息	机构统一社会信用代码
003001	证照信息类	自然人证照信息	身份证件号码
003002		机构证照信息	机构统一社会信用代码
004001	交通运输类	交通工具类信息	身份证件号码、发动机编号
004002		交通事件类信息	主体身份证件号码
004003		交通设备类信息	设备编号
005001	医疗卫生类	患者信息	身份证件号码
005002		医疗卫生类事件信息	事件编号
006001	安全生产类	安全生产人员类信息	身份证件号码
006002		安全生产事件类信息	事件编号
007001	信用服务类	行政处罚类信息	统一社会信用代码
007002		信用评价类信息	统一社会信用代码
008001	社保就业类	社保就业人员类信息	身份证件号码
008002		参保机构类信息	统一社会信用代码
008003		社保就业类案件信息	对象统一社会信用代码
009001	城建住房类	建筑信息	身份证件号码、建筑物名称
009002		证件信息	身份证件号码
009003		公积金信息	身份证件号码

009004		城建住房类项目信息	项目编号
010001	教育文化类	教职工人员信息	身份证件号码
010002		学生类人员信息	身份证件号码
011001	生态环境类	生态区域类信息	区域名称
011002		生态环境类事件信息	违法单位
011003		监测指标类信息	指标名称
012001	科技创新类	成果信息	成果名称、成果完成人
013001	财税金融类	纳税人信息	统一社会信用代码
013002		经济指标信息	指标名称、税务机关名称
014001	工业制造类	项目类信息	项目代码

表 A.1 信息类别及关键索引数据元表（续）

标识符	信息类别	关键索引数据元
014002	工业制造类	工业制造类经济主体信息
014003		工程类信息
014004		工业制造产品类信息
014005		工业制造类指标信息
015001	农业农村类	农业类项目信息
015002		农业类经济主体信息
015003		农村类信息
015004		农业产品类信息
016001	商贸服务类	商贸服务类信息
016002		公共服务类信息
017001	市场监管类	市场主体信息
017002		场所信息
017003		处罚信息
018001	公共安全类	安全事件类信息
018002		公共设施信息
019001	资源能源类	资源类信息
019002		交易信息
020001	空间地址类	地址类信息
020002		空间地理类信息

## 附录 B

(资料性)

### 核心业务数据元

B.1 核心业务数据元（见表 B.1）通过以下属性进行描述：

- 标识符，
- 信息类别，
- 名称，
- 定义，
- 数据类型，
- 数据格式，
- 备注，
- 数据项级别，
- 数据集级别。

B.2 标识符通过在信息类别标识符（见附录 A）基础上增加一段三位编码来标识，该三位数字为同一信息类别中数据元的顺序号。

示例：“兵役状况”的标识符是 001001006，其中前两段各三位“001001”为该数据元所属信息类别的标识，后三位“006”为该数据在自然人基本信息数据元中的顺序号。

B.3 数据级别按 DB33/T 2351-2021 的分级方法和判定标准确定，并以信息类别下属数据项的最高数据级别为数据集级别。

表 B.1 核心业务数据元

标识符	信息类别		名称	定义	数据类型	数据格式	备注	数据项级别	数据集级别
001001001	自然人基 本信息类	自然人基 本信息	姓名	在户籍管理部门正式登记注册、人事档案中正式记载的姓名名称	字符型	C..50		L1	L3
001001002			性别	人的基本生理特征	字符型	C..20		L1	
001001003			出生日期	出生证签署的，并在户籍部门正式登记注册、人事档案中记载的日期	日期型	YYYYMMDD		L2	
001001004			身份证件号码	身份证件上记载的、可唯一标识个人身份的号码	字符型	C..18		L3	

表 B.1 核心业务数据元（续）

标识符	信息类别		名称	定义	数据类型	数据格式	备注	数据项级别	数据集级别	
001001005		自然人基本信息	兵役状况	是否服兵役	字符型	C..20		L2	L3	
001001006			户籍地址	身份证登记的户口所在地地址	字符型	C..100		L3		
001001007			居住地址	公民连续居住一定合理期限的地方	字符型	C..40		L3		
001002001	自然人基本信息类	自然人职业信息	姓名	在户籍管理部门正式登记注册、人事档案中正式记载的姓名名称	字符型	C..50		L1	L3	
001002002			性别	人的基本生理特征	字符型	C..20		L1		
001002003			身份证件号码	身份证件上记载的、可唯一标识个人身份的号码	字符型	C..18		L3		
001002004			职业	个人服务社会并作为主要生活来源的工作名称	字符型	C..50		L2		
001002005			工作单位名称	人员目前供职的工作单位名称	字符型	C..100		L2		
001002006			工作单位统一社会信用代码	用代码	一组长度为 18 位的用于法人和其他组织身份识别的唯一的代码	字符型	C18			L2
002001001			统一社会信用代码	一组长度为 18 位的用于法人和其他组织身份识别的唯一的代码	字符型	C18		L1		

	机构团体 类	机构基本 信息							L1
002001002			机构名称	能够反映出该机构的行政区划、所属关系、工作性质、规格级别以及管 理范围等	字符型	C. . 100		L1	
002001003			机构类型	以企业登记机关、机构编制管理机关核定或确定的类型为准，参照其它 法律、法规规定的组织机构登记或批准机关核定或确定的类型	字符型	C. . 20		L1	
002001004			成立日期	机构注册或被批准成立的日期	日期型	YYYYMMDD		L1	
002001005			法定代表人	机构登记机关或批准机关核发的有效证照或批文上的法定代表人的姓名	字符型	C. . 50		L1	
002001006			机构地址	机构进行具体生产、经营、服务活动的地址信息	字符型	C. . 100		L1	

表 B.1 核心业务数据元（续）

标识符	信息类别	名称	定义	数据类型	数据格式	备注	数据项级别	数据集级别
003001001	自然人证照信息	证照名称	在户籍管理部门正式登记注册、人事档案中正式记载的姓名名称	字符型	C..50		L3	L3

	证照信息 类								
003001002		证照编号	人的基本生理特征	字符型	C. . 20			L3	
003001003		证照持有人姓名	出生证签署的，并在户籍部门正式登记注册、人事档案中记载的日期	日期型	YYYYMMDD			L1	
003001004		身份证件号码	身份证件上记载的、可唯一标识个人身份的号码	字符型	C. . 18			L3	
003001005		证照颁发机构	与照面信息一致的颁发证照的机构名称	字符型	C. . 40			L2	
003001006		证照颁发时间	证照颁发的时间	日期型	YYYYMMDD			L2	
003001007		证照有效期截止日期	证照有效时间的最后一天日期	日期型	YYYYMMDD	无有效期 限以特定 字符代替		L2	
003002001		证照名称	与照面信息一致的证照名称	字符型	C. . 40			L2	
	机构证照 信息								L2
003002002		证照颁发机构	与照面信息一致的颁发证照的机构名称	字符型	C. . 40			L2	
003002003		证照颁发时间	证照颁发的时间	日期型	YYYYMMDD			L2	
003002004		证照有效期截止日期	证照有效时间的最后一天日期	日期型	YYYYMMDD	无有效期 限以特定 字符代替		L2	
003002005		机构名称	能够反映出该机构的行政区划、所属关系、工作性质、规格级别以及管	字符型	C. . 100			L2	

			理范围等				
003002006		机构法定代表人	机构登记机关或批准机关核发的有效证照或批文上的法定代表人的姓名	字符型	C. . 50		L2
003002007		统一社会信用代码	一组长度为 18 位的用于法人和其他组织身份识别的唯一的代码	字符型	C18		L1

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/066042151052011004>