

ICS

CCS 点击此处添加 CCS 号

DB51

四川省地方标准

DB51/T XXXX—XXXX

公共数据资源管理
第2部分：数据资源归集

点击此处添加标准名称的英文译名

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

四川省市场监督管理局 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 基本原则 1

5 数据归集流程 2

6 明确归集范围 2

7 明确归集方式 2

8 开展数据归集 2

9 归集结果确认 6

10 数据更新 6

11 数据归集安全 7

12 数据销毁 7

附录 A（资料性） 库表归集相关数据表约束及示例 8

附录 B（资料性） 文件归集相关数据表约束及示例 10

附录 C（资料性） 服务接口数据示例及内容描述 11

附录 D（资料性） 数据资源汇聚清单模板 14

参考文献 15

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。

DB51/T XXXX《公共数据资源管理》分为3个部分：

——第1部分：数据资源目录编制；

——第2部分：数据资源归集；

——第3部分：数据资源登记；

本文件为《公共数据资源管理》的第2部分。

本文件由四川省大数据中心提出并归口。

本文件起草单位：

本文件主要起草人：

公共数据资源管理

第2部分：数据资源归集

1 范围

本文件规定了公共数据的归集流程、归集范围、归集方式、归集结果确认、数据更新等的要求。
本文件适用于指导四川省行政区域范围内公共数据的归集。

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 37973-2019 信息安全技术 大数据安全管理指南

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

国家机关和法律、法规授权的具有管理公共事务职能的组织，以及医疗、教育、供水、供电、供气、通信、文化旅游、体育、交通运输、环境保护等公共企业事业单位。

3.2

数据管理机构 public data administration institution

各级人民政府明确的负责本行政区域数据管理、开发利用和监督检查，以及推动数据资源体系建设和数据要素市场培育的机构。

3.3

数据归集 data collection and aggregation

数据管理机构面向公共数据提供方，对公共数据进行收集、汇聚、整合的行为和过程。

3.4

省一体化数据管理系统 Provincial Integrated Data Management System

全省公共数据管理、共享开放的统一基础支撑平台，由省级枢纽、市级节点和县级节点组成。

4 基本原则

4.1 一致性

数据提供方需按照数据汇聚清单确认的方式进行数据归集。

注：数据汇聚清单详见附录D数据资源汇聚清单模板。

4.2 及时性

数据提供方需按照数据汇聚清单确认的频次，按时提供增量数据。

4.3 真实性

数据提供方应保证所提供数据的真实性，确保所提供数据无篡改、无遗漏。

5 数据归集流程

5.1 流程概述

公共数据归集包括明确归集范围、明确归集方式、开展数据归集、归集结果确认四个阶段。

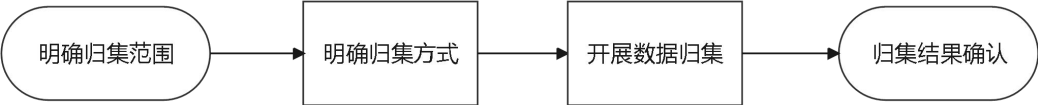


图 1 数据归集流程图

6 明确归集范围

公共管理和服务机构应根据业务职责，按照《公共数据资源管理第1部分：数据资源目录编制》编制并动态维护公共数据资源目录，按照“按需归集、应归尽归”的原则，将本机构所有可归集公共数据持续归集至省一体化数据管理系统。

7 明确归集方式

基于省一体化数据管理系统底座能力，数据归集方式有以下三种方式：

- a) 库表方式，适用于各类结构化数据的归集；
- b) 文件方式，适用于各类非结构化数据的归集。涉及国家垂直管理系统中的数据，数据提供方经申请后导出的离线文件也可采用文件方式归集；
- c) 接口方式，适用于统计指标类数据的归集。部分数据提供方的数据因具有敏感性，无法提供明细数据。这类数据可采用接口方式进行归集，其接口要求支持不低于分钟级查询。

8 开展数据归集

8.1 以库表方式归集

8.1.1 主要工作

8.1.1.1 应在前置机中建设前置库，用于临时存储业务数据表、对账数据表等，可根据业务需要创建多个前置库。

8.1.1.2 应在前置库中建设业务数据表、对账数据表，业务数据表存储待归集的业务数据，对账数据表存储数据的更新情况。

注：前置机主要部署在公共管理和服务机构端，用于实现库表、文件数据的批量传输。

8.1.2 归集要求

8.1.2.1 前置库

前置库应支持国产数据库，字符编码应为UTF-8。

8.1.2.2 业务数据表

业务数据表除业务字段外应包含记录ID、批次号、同步时间、同步类型等扩展信息。业务数据表扩展信息的字段名称和说明见表1，业务数据表的约束和示例见附录A.1。

表 1 业务数据表扩展信息

| 字段名称 | 字段说明 |
|-------|--|
| 记录 ID | 记录 ID 为业务数据主键，是业务数据的唯一记录标识。 |
| 批次号 | 批次号为业务数据按批次更新时生成，是数据日期和数据序号的组合，应保证唯一性。其中数据日期指数据更新的日期，格式为 YYYYMMDD；数据序号为 8 位数字，每日从 00000001 顺序递增。以 2022051000000003 为例，该数据序号表示 2022 年 5 月 10 日第 3 批次数据。 |
| 同步时间 | 更新时间为业务数据写入业务数据表的时间，格式为 YYYY-MM-DD hh:mm:ss。 |
| 同步类型 | 业务操作标识用于标识业务数据的操作属性，数据新增为“I”，数据修改为“U”，数据删除为“D”。 |

8.1.2.3 对账数据表

对账数据表见表2，约束和示例见附录A.2。

表 2 对账数据表

| 字段名称 | 说明 |
|----------|---|
| 对账记录唯一标识 | 对账表主键 |
| 批次号 | 批次号为业务数据按批次更新时生成，是数据日期和数据序号的组合，应保证唯一性。其中数据日期指数据更新的日期，格式为 YYYYMMDD；数据序号为 8 位数字，每日从 00000001 顺序递增。以 2024051400000002 为例，该数据序号表示 2024 年 5 月 14 日第 2 批次数据 |
| 批次时间 | 对账数据写入对账表的时间，采用数据库时间函数的值，格式为 YYYY-MM-DD hh:mm:ss |
| 批次数据条数 | 该批次更新的数据条数，应与该批次结构化文件中数据实际更新条数一致 |
| 批次状态 | 本批次数据归集状态，0 为初始默认值，1 为完成归集，-1 为对账表批次条数或文件个数与实际不一致。 |

8.2 以文件方式归集

8.2.1 主要工作

- 8.2.1.1 应在前置机中建设文件系统，用于临时存储待归集的文件。
- 8.2.1.2 在前置库中建设文件信息表、文件拆分信息表、对账数据表，文件信息表、文件拆分信息表分别存储非结构化文件的详细信息和拆分信息，对账数据表存储文件的更新情况。

8.2.2 归集要求

8.2.2.1 提供方式

文件系统可采用离线拷贝、FTP、SFTP等。

8.2.2.2 文件命名

数据文件命名应由数据提供部门名称首字母、数据表名、下划线“_”、数据提供日期及时间（精确到小时）、两位阿拉伯数字组成。命名格式为：scsdsjzx_ryxx_2025071109__01，两位阿拉伯数字用于标记对同一个大文件拆分情况。

8.2.2.3 文件大小

应设置文件大小阈值，文件超出阈值应拆分成多个小文件。

8.2.2.4 文件格式

结构化文件格式只限于CSV、TXT、XLS、XLSX四种格式。数据文件首行为列名，字符编码应为UTF-8，且不存在特殊字符，其他约束包括但不限于：

- a) TXT、CSV 文件格式的数据以半角逗号作分隔符；
- b) XLS、XLSX 文件格式的约束主要包括：应只包括一个表单页；不应存在行或列的合并；不应存在相同列名；不应有空白行。

8.2.2.5 文件内容

结构化文件中除业务字段外，应包含记录ID、批次号、同步时间、同步类型等扩展字段信息。

8.2.2.6 文件信息表

文件信息表见表3，约束和示例见附录B.1。

表 3 文件信息表

| 字段名称 | 说明 |
|---------|---|
| 批次号 | 批次号为业务数据按批次更新时生成，是数据日期和数据序号的组合，应保证唯一性。其中数据日期指数据更新的日期，格式为 YYYYMMDD；数据序号为 8 位数字，每日从 00000001 顺序递增。以 2024051400000002 为例，该数据序号表示 2024 年 5 月 14 日第 2 批次数据 |
| 文件个数 | 该批次的文件的个数，应与该批次非结构化文件的实际个数相符 |
| 数据表数量 | 示例：XX 张 |
| 数据量 | 示例： 数据表 1 名称：XX 条 数据表 2 名称：XX 条 |
| 是否包含拆分表 | 是 / 否 |

8.2.2.7 对账数据表

文件的对账数据表应包括对账记录唯一标识、文件路径、批次号、批次时间、批次状态等字段，其中结构化文件对账数据表还应包括批次数据条数等字段，字段标识符以“SD”开头。对账数据表见表4，约束和示例见B.1。

表 4 对账数据表

| 字段名称 | 字段说明 |
|----------|---|
| 对账记录唯一标识 | 对账记录唯一标识为对账表主键，宜使用 UUID 主键策略。 |
| 文件路径 | 文件路径为拆分文件的存储路径。 |
| 批次号 | 批次号为业务数据按批次更新时生成，是数据日期和数据序号的组合，应保证唯一性。其中数据日期指数据更新的日期，格式为 YYYYMMDD；数据序号为 8 位数字，每日从 00000001 顺序递增以 2024051400000002 为例，该数据序号表示 2024 年 5 月 14 日第 2 批次数据。 |
| 批次时间 | 批次时间为对账数据写入对账表的时间，采用数据库时间函数的值，格式为 YYYY-MM-DD hh:mm:ss。 |
| 批次数据条数 | 批次条数为该批次更新的数据条数，应与该批次结构化文件中数据实际更新条数一致。 |
| 批次个数 | 批次个数为该批次的文件的个数，应与该批次非结构化文件的实际个数相符。 |
| 批次大小 | 批次大小为该批次的文件占用的存储空间，应与该批次文件实际大小相符，单位为 KB。 |
| 批次状态 | 批次状态为本批次数据归集状态，0 为初始默认值，1 为完成归集，-1 为对账表批次条数或文件个数与实际不一致。 |

8.3 以服务接口方式归集

8.3.1 主要工作

- 8.3.1.1 数据管理机构应根据提供的库表结构开发服务接口，挂载到相关公共数据资源目录下，通过省一体化数据管理系统发布。
- 8.3.1.2 应依据服务接口要求，开发服务接口调用程序，调用服务接口推送本机构数据至省一体化数据管理系统，可采用压缩等技术实现数据高速传输。

8.3.2 归集要求

8.3.2.1 接口调用形式

接口调用形式为以下四种：

- a) Protocol 支持 HTTP、HTTPS；
- c) Hostname 指 IP 设备的 IP 地址；
- d) Port 指端口号；
- e) URI 指资源 URI，由接口唯一标识、数据传输流水序列号、分布式存储标识、资源唯一标识符组成，具体构成为：
`<IneterIdentifier>/<serialnumber>/<DistributedStorageIdentifier>/<ResouceIdentifier>`。

8.3.2.2 服务接口资源描述

接口资源采用Schema架构说明的标准XML文件方式进行描述，其中编码方式为UTF-8，XMLSchema描述了内容传输格式文档的结构，内容传输格式应根据公共数据资源目录规定的格式类型等属性确定。XML格式数据示例见附录C. 1。

8.3.2.3 接口功能

接口主要分为增删改类数据服务接口和只增类数据服务接口，服务接口的功能见表5，接口内容描述见附录C. 2~C. 5。

表 5 接口功能

| 序号 | 分类 | 功能 |
|----|------------|--------|
| 1 | 增删改类数据服务接口 | 增加信息 |
| 2 | | 增加压缩信息 |
| 3 | | 删除信息 |
| 4 | | 删除压缩信息 |
| 5 | | 修改信息 |
| 6 | | 修改压缩信息 |
| 7 | 只增类数据服务接口 | 只增信息 |
| 8 | | 只增压缩信息 |

9 归集结果确认

由数据主管单位对归集的过程和结果进行核查确认，不符合要求的由数据提供部门进行整改。

10 数据更新

10.1 更新方式

更新方式需与归集方式相同。

10.2 更新频率

更新频率见表6。

表 6 更新频率

| 数据类型 | 更新周期 | 归集频率 |
|------|------|----------|
| 实时类 | 分钟级 | ≤ 10 min |
| 每日业务 | 天级 | T+1 |
| 每周统计 | 周级 | 每周周二前 |
| 月度统计 | 月级 | 次月 5 日前 |
| 年度统计 | 年级 | 次年1月前 |
| 档案类 | 不定期 | 按需 |

10.3 变更记录

确保应收数据量与实收数据量一致，数据提供方应配合开展数据对账工作。

11 数据归集安全

数据归集安全应符合 GB/T 22239-2019 中等级保护三级的要求，个人信息安全应符合 GB/T 35273-2020 要求，其他安全要求应符合 GB/T 37973-2019 的要求。

12 数据销毁

综合考虑数据安全需要，数据管理机构应与数据提供方沟通一致后，定期清理销毁进入前置机超过 30 日的的数据。

附录 A
(资料性)
库表归集相关数据表约束及示例

A.1 业务数据表约束及示例

A.1.1 业务数据表约束

业务数据表包括记录ID、批次号、业务操作标识、更新时间以及多个业务字段，除业务字段外相关约束见表A.1。

表 A.1 业务数据表约束

| 字段名称 | 字段标识符 | 字段类型 | 字段长度 | 主键标识 | 非空 |
|-------|--------------|----------|------|------|----|
| 记录 ID | ID | VARCHAR | 64 | 主键 | 是 |
| 批次号 | CD_BATCH | VARCHAR | 20 | — | 是 |
| 同步类型 | CD_OPERATION | VARCHAR | 1 | — | 是 |
| 同步时间 | CD_TIME | DATETIME | 8 | — | 是 |

A.1.2 业务数据表示例

业务数据表示例见表A.2。

表 A.2 业务数据表示例

| ID | 业务字段1 | ... | 业务字段N | CD_BATCH | CD_OPERATION | CD_TIME |
|------|-------|-----|-------|------------------|--------------|------------------------|
| UUID | 业务字段 | ... | 业务字段 | 2024071400000001 | I | 2024-07-14 17:20:53 |
| UUID | 业务字段 | ... | 业务字段 | 2024071400000001 | U | 2024-07-14 17:30:53 |
| UUID | 业务字段 | ... | 业务字段 | 2024071400000001 | D | 2024-07-14 17:40:53 |
| UUID | 业务字段 | ... | 业务字段 | 2024071400000002 | I | 2024-07-14 17:20:53 |

A.2 对账数据表约束及示例

A.2.1 对账数据表约束

对账数据表包括对账记录唯一标识、表名、批次号、批次数据条数、批次时间、批次状态等字段，相关约束见表A.3。

表 A. 3 对账数据表约束

| 字段名称 | 字段标识符 | 字段类型 | 字段长度 | 主键标识 | 非空 |
|----------|--------------|----------|------|------|----|
| 对账记录唯一标识 | ID | VARCHAR | 64 | 主键 | 是 |
| 表名 | TABLE_NAME | VARCHAR | 100 | — | 是 |
| 批次号 | CD_OPERATION | VARCHAR | 20 | — | 是 |
| 批次数据条数 | COUNT | INT | — | — | 是 |
| 批次时间 | DATA | DATETIME | 8 | — | 是 |
| 批次状态 | STATE | VARCHAR | 1 | — | 是 |

A. 2. 2 对账数据表约束

对账数据表示例见A. 4

表 A. 4 对账数据表示例

| ID | TABLE_NAMEBATCH | COUNT | DATA | STATE |
|------|------------------|-------|------------------------|-------|
| UUID | 2024071400000001 | 1000 | 2024-07-14 17:20:53 | 1 |

附录 B
(资料性)
文件归集相关数据表约束及示例

B.1 对账数据表约束及示例

B.1.1 对账数据表约束

结构化文件的对账数据表应包括对账记录唯一标识、文件路径、批次号、批次条数、批次时间、批次状态等字段，相关约束见表B.1。

表 B.1 结构化文件对账数据表约束

| 字段名称 | 字段标识符 | 字段类型 | 字段长度 | 主键标识 | 非空 |
|----------|--------------|----------|------|------|----|
| 对账记录唯一标识 | SD_ID | VARCHAR | 64 | 主键 | 是 |
| 文件路径 | SD_FILE_PATH | VARCHAR | 250 | — | 是 |
| 批次号 | SD_BATCH | VARCHAR | 20 | — | 是 |
| 批次条数 | SD_COUNT | INT | — | — | 是 |
| 批次时间 | SD_DATA | DATETIME | 8 | — | 是 |
| 批次状态 | SD_STATE | VARCHAR | 1 | — | 是 |

B.1.2 对账数据表示例

结构化文件的对账数据表示例见表B.2。

表 B.2 结构化文件的对账数据表示例

| SD_ID | SD_FILE_PATH | SD_BATCH | SD_COUNT | SD_DATA |
|-------|--------------------|------------------|----------|------------------------|
| UUID | /SD_SJGJ_SOURCE... | 2024071400000001 | 1000 | 2024-07-14 17:20:53 |

附录 C

(资料性)

服务接口数据示例及内容描述

C.1 XML 格式数据示例

以test内容传输格式为例，定义数据模板，Schema验证模板标识ID为test，信息表名称为wci，XIMIL编码为UTF-8，XL格式数据示例为：

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<test>
  xmlns=""
  <xs:schema
    ID="test"
    xmlns:xs="http://www.w3.org/2001/XMLSchema
    xmlns:msdata="urn:schemas-microsoft-com:xml-msdata">
    <xs:element name="test" msdata:IsDataSet="true" msdata:UseCurrentLocale="true">
    <xs:complexType>
    <xs:choice minOccurs="0" maxOccurs="unbounded">
    <xs:element name="wci">
    <xs:complexType>
    <xs:sequence>
    <xs:element name="ApeID" type="xs:string" minOccurs="0" />
    <xs:element name="mac" type="xs:string" minOccurs="0" />
    <xs:element name="zdcq" type="xs:string" minOccurs="0" />
    <xs:element name="rdssID" type="xs:string" minOccurs="0" />
    <xs:element name="cjsj" type="xs:dateTime" minOccurs="0" />
    </xs:sequence>
    </xs:complexType>
    </xs:element>
    </xs:choice>
    </xs:complexType>
    </xs:element>
    </xs:schema>
  <wci>
    <ApeID>abcdefghij1234567890</ApeID>
    <mac>12345543211234567890</mac>
    <zdcq>123456.44</zdcq>
    <rdssID>sdsgat</rdssID>
    <cjsj>2019-11-06T08:09:58.99+08:00</cjsj>
    </wci>
  </test>
```

C.2 增加信息接口

增加信息服务接口内容描述见表C.1

表 C.1 增加信息服务接口内容描述

| 名称 | 内容 | 备注 |
|------|--|--|
| 接口地址 | https://<Ilostname>:<PortXIneterfacelIdentifier>/serialnumb er>/<DistributedStoragelIdentifier>/<ResouceIdentifier> | IneterfacelIdentifier 为 addob jects或addob jects GZIP |
| 功能 | 批量增加信息 | — |
| 方法 | post | — |
| 消息体 | [neterfacelIdentifier=addobjects时消息体为UTF-8编码带schema 架构说明的基本数据类型标准xml; [neterfaceIdentifier=addobjects GZIP时消息体为UTF-8 编 码 带 schema架构说明的基本数据类型标准xml的GZIP压缩包 | — |
| 返回结果 | <?xml version="1.0" encoding="utf-8" standalone="yes"?><response status "{status}" err="{err}" /> | {status}=0, 数据格式正确, {err}为空 {status}不为0, 数据 格式错误, {err}为错误描述 |
| 注释 | — | 需主键信息 |

C.3 删除信息接口

删除信息服务接口内容描述见表C.2

表 C.2 删除信息服务接口内容描述

| 名称 | 内容 | 备注 |
|------|---|---|
| 接口地址 | https://<Hostname>:<Port><IneterfacelIdentifier>/<serialnumb er>/<DistributedStoragelIdentifier>/<ResouceIdentifier> | IneterfacelIdentifier为 deleteobjects、deleteobjects GZIP |
| 功能 | 批量删除信息 | — |
| 方法 | post | — |
| 消息体 | IneterfaceIdentifier=deleteob. jects时消息体为UTF-8编码带 schema架构说明的基本数据类型标准xml; IneterfaceIdentifier=deleteob iects GZIP时消息体为UTF-8编码带 schema架构说明的基本数据类型标准xml的GZIP压缩包 | — |
| 返回结果 | <?xml version="1.0" encoding="utf-8" standalone="yes"?><response status "{status}" err="{ferr}" /> | {status}=0, 数据格式正确, {err}为空: {status}不为0, 数据格式错误, {err}为错误描述 |
| 注释 | — | 需主键信息 |

C.4 修改信息接口

修改信息服务接口内容描述见 表C.3

表 C.3 修改信息服务接口内容描述

| 名称 | 内容 | 备注 |
|------|--|---|
| 接口地址 | https://<Hostname>:<Port><IneterfaceIdentifier>/<serialn umber>/DistributedStorageIdentifier>/<ResouceIdentifier | IneterfaceIdentifier 为 updateobjects.updateobjects GZIP |
| 功能 | 批量修改信息 | — |
| 方法 | post | — |
| 消息体 | IneterfaceIdentifier=updateobjects时消息体为UTF-8编码带 schema架构说明的基本数据类型标准xml; IneterfaceIdentifier=updateobjects GZIP时消息体为UTF-8编码带 schema架构说明的基本数据类型标准xml的GZIP压缩包 | — |
| 返回结果 | <?xml version="1.0"encoding="utf-8 "standalone="yes"?><response status="{status}err="{err}" /> | {status}=0, 数据格式正确, {err}为空: status不为0, 数据格式错误, {err}为错误描述 |
| 注释 | — | 需主键信息 |

C.5 只增信息接口

只增信息服务接口内容描述见表C.4

表 C.4 只增信息服务接口内容描述

| 名称 | 内容 | 备注 |
|------|--|---|
| 接口地址 | https://<Hostname>:<Port><IneterfaceIdentifier>/<serialn umber>/DistributedStorageIdentifier>/<ResouceIdentifier | IneterfaceIdentifier 为addeventinfos. addeventinfos GZIP |
| 功能 | 批量归集只增信息 | — |
| 方法 | post | — |
| 消息体 | IneterfaceIdentifier=addeventinfos时消息体为UTF-8编码带schema 架构说明的基本数据类型标准xml; IneterfaceIdentifier=addeventinfos GZIP时消息体为UTF-8编码带 schema架构说明的基本数据类型标准xml的GZIP压缩包 | — |
| 返回结果 | <?xml version="1.0"encoding="utf-8" standalone="yes"?><response status="{status}err="{lerr}" /> | {status}=0, 数据格式正确, {err}为空; {status}不为0, 数据格式错误, {err}为错误描述 |
| 注释 | — | 需主键信息 |

附 录 D
(资料性)
数据资源汇聚清单模板
XX厅数据资源汇聚清单

| 序 号 | 基 础 库 类 型 | 数据资 源表名 称 | 数据项 名称 | 提供方式 | 开始汇聚时 间 | 完成汇聚 时间 | 责任人及电 话（厅级） | 联络人及电 话（处级） | 后续更 新频率 | 备注 |
|--------|-----------------------|-----------------|-----------|------|------------|------------|----------------|----------------|------------|----|
|--------|-----------------------|-----------------|-----------|------|------------|------------|----------------|----------------|------------|----|

参 考 文 献

- [1] GB/T 33850-2017 信息技术服务质量评价指标体系
 - [2] GB/T 39477-2020 信息安全技术—政务信息共享—数据安全技术要求
-