ICS 35.020
CCS L 72

DB31

上　海　市　地　方　标　准

DB 31/T XXXX—XXXX

大数据资源平台共享交换要求

Requirements for the sharing and exchange of the data resource platform

（征求意见稿）

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

上海市市场监督管理局　　发 布

目　次

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。

本文件由上海市数据局提出并组织实施。

本文件由上海市数据标准化技术委员会归口。

本文件起草单位：上海市数据局、上海市大数据中心、云赛智联股份有限公司、上海数据集团有限公司。

本文件主要起草人：。

大数据资源平台共享交换要求

1 范围

本文件规定了大数据资源平台数据共享交换方式、数据归集、数据共享、数据安全等要求。

本文件适用于依托大数据资源平台开展的公共数据共享交换工作

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 21062 政务信息资源交换体系

GB/T 22239-2019 信息安全技术 网络安全等级保护基本要求

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 43697 数据安全技术 数据分类分级规则

DB 31/T XXXX 长三角数据共享交换平台 数据接入规范

DB31/T 1241-XXXX 公共数据“三清单”管理规范

DB31/T 1523 公共数据质量评价要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

大数据资源平台 the data resource platform

依托电子政务云实施全市公共数据归集、整合、共享、开放、运营的统一基础设施。

3.2

数据湖 data lake

大数据资源平台的组成部分，集中管理全市跨系统、跨行业或跨领域公共数据的一种高度可扩展的数据存储架构。

3.3

数据池 data pool

大数据资源平台的组成部分，汇聚、治理、融合和共享单个系统、行业或领域数据的一种高度可扩展的数据存储架构。

3.4

目录链 catalog chain

大数据资源平台的组成部分，是指利用区块链技术，对本市公共数据目录进行统一管控的分布式系统。

4 概述

大数据资源平台具备数据归集、数据治理分析基础能力,通过前置交换、平台级联等方式提供数据共享交换服务，形成覆盖国省、长三角区域、市本级、市区等各层级的数据共享交换体系：

a） 市级公共管理和服务机构依托目录链，通过库表、文件、服务接口、消息队列等交换方式，实现国家、市级、区级数据共享交换；

b） 区大数据资源分平台与大数据资源平台以市区级联方式实现双向数据访问；

c） 大数据资源平台与国家数据共享交换平台、长三角数据共享交换平台以平台级联方式完成对接，实现跨层级、跨区域、跨部门、跨系统的数据访问。

大数据资源平台数据共享交换架构见图 1。

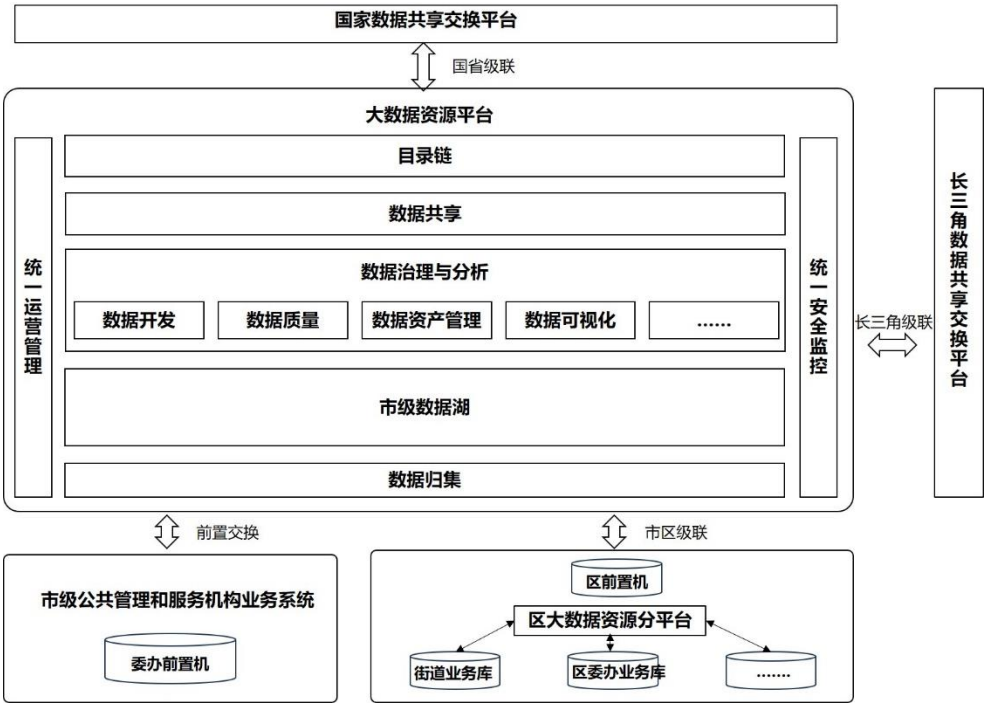

图1　大数据资源平台数据共享交换架构

5　交换方式

5.1　库表交换

库表交换依托前置机实现不同数据库，或同一数据库内不同表之间的数据、表结构或元数据的转移与同步，适用于适合数据量较大且需要落地的公共数据共享。

5.2　文件交换

文件交换依托前置机在在不同设备、系统或用户之间传输、共享或同步文件数据，适用于实时性要求较低，数据量大的非结构或半结构化数据，如多媒体数据。

5.3　服务接口交换

服务接口交换是通过定制化开发接口的方式提供数据共享服务，适用于业务协同、信息核验等实时性要求高的数据应用场景。

5.4 消息队列交换

消息流交换是以消息队列承载数据同步，实现消息流在数据提供方和数据使用方之间传递的共享交换方式，适用于数据实时共享交换场景。

6 数据归集

6.1 归集策略

数据归集包括定时归集、不定时归集、实时归集：

a) 定时归集：按照固定的周期性间隔，通过对归集任务自动化调度执行的归集方式，主要采用每天、每周、每月、每季度、每半年、每年等归集策略；
b) 不定时归集：一次性或者临时性归集，常用于初始数据采集，临时数据更新和异常数据修补；
c) 实时归集:数据产生、传输、处理和存储过程中，实时将分散在不同来源的数据快速收集、整合。

6.2 库表归集

6.2.1 归集步骤

库表数据归集步骤如下：

a) 数据提供方按照 DB31/T 1523 相关要求和数据产生的业务规则，对数据的规范性、完整性、准确性、一致性、时效性、可访问性等进行质量检查；
b) 数据提供方按照业务需求推送数据到前置库；
c) 数据提供方通过目录链管理系统，将所需归集的数据注册编目，并发起数据归集任务；
d) 使用对账表形式进行前置库对账，对账信息包括库名、表名、库类型、对账结果、数据量、交换时间等；
e) 大数据资源平台通过平台归集节点从前置库抽取数据，并写入市级数据湖；
f) 针对大数据资源平台的数据质量平台或由数据使用方反馈的异议数据，质量稽核运营团队发起对账异常工单并与数据提供方进行数据量核实。经过异议核实处理后，重新归集数据；
g) 数据提供方按数据更新周期持续推送增量数据，大数据资源平台根据规定的数据归集时间进行抽取。如无增量数据产生，数据提供方应在对账表中填报“0”。

6.2.2 技术要求

库表归集相关技术要求如下：

a) 归集频率：分钟、小时、天、月、年；
b) 归集类型：全量采集、增量采集；
c) 前置库类型：支持主流关系型数据库的库表归集。

6.2.3 操作要求

前置库在初始化和增量归集库表过程中应符合以下要求：

a) 初始化时，添加数据库归集时间戳字段，字段类型为 TimeStamp，命名为：jhpt_update_time；
b) 初始化时，添加数据库删除标识字段，字段类型为 Int(1)，命名为 jhpt_delete；
c) 初始化时，库表命名及字段命名长度控制在 128 位以内，浮点数字段类型长度控制在 38 位以内；
d) 初始化时，添加主键字段；
e) 增量归集过程中，数据提供方将本次推入的数据量同步至对账表，填写相关字段信息，包括表名、统计开始时间、统计结束时间、统计时间范围内数据总条数等；
f) 当数据结构发生变更时，对原有的数据资源下线，并将变更后的数据资源重新申请发布，经审批后，再重新发起归集任务。

6.3 文件归集

6.3.1 归集步骤

文件归集步骤如下：

a) 数据提供方按照 DB31/T 1523 相关要求和数据产生的业务规则，对数据的规范性、完整性、准确性、一致性、时效性、可访问性等进行质量检查；
b) 数据提供方按照业务需求上传文件至前置机；
c) 数据提供方通过目录链管理系统，将所需归集的数据注册编目，并发起数据归集任务；
d) 对归集的文件标注数据量条数和该文件的 md5 值，以确保该文件入湖对账的一致性；
e) 大数据资源平台通过平台归集节点进行抽取，并对具有固定结构的文件解析、写入市级数据湖；
f) 针对大数据资源平台的数据质量平台或由数据使用方反馈的异议数据，质量稽核运营团队发起对账异常工单并与数据提供方进行数据量核实。经过异议核实处理后，重新归集数据；
g) 数据提供方按数据更新周期持续推送增量数据，大数据资源平台根据规定的数据归集时间进行抽取。如无增量数据产生，数据提供方应在对账表中填报“0”。

6.3.2 技术要求

文件归集相关技术要求如下：

a) 归集频率：
   1) 半结构化文件：分钟、小时、天、月、年；
   2) 非结构化文件：天、月。
b) 归集类型：
   1) 半结构化文件：增量采集；
   2) 非结构化文件：全量采集、增量采集。
c) 文件类型：支持 txt、csv、xml 等格式。

6.3.3 操作要求

6.3.3.1 前置库在初始化和增量归集结构化文件过程中应符合下列要求：

a) 初始化时，添加数据库归集时间戳字段，字段类型为 TimeStamp，命名为：jhpt_update_time；

b) 初始化时，添加数据库文件路径字段，字段类型为 Int(1)，命名为 jhpt_delete；
c) 文本文件内容编码使用 utf-8 编码；
d) 文件内容的字段顺序与编目字段顺序保持一致；
e) txt 文件分隔符为‘\u0001’隐藏字符；
f) csv 文件为‘,’分隔，字段数据中不应出现英文逗号、回车换行符等文本内容；
g) excel 文件首行为表头，不支持多工作表（sheet）采集；
h) 前置机节点在数据文件目录上传与数据文件同名的对账文件。

6.3.3.2 前置库在初始化和增量归集非结构化文件过程中应符合下列要求：

a) 初始化时，添加数据库归集时间戳字段，字段类型为 TimeStamp，命名为：jhpt_update_time；
b) 初始化时，添加数据库文件路径字段，字段类型为 Int(1)，命名为 jhpt_delete；
c) 前置机数据库中，必须有主键字段并且 jhpt_file_path 字段的内容为 FTP/SFTP 的根路径，且附件字段与 FTP/SFTP 服务器上的文件一定要对应，不允许为空，必须以‘/’开头；
d) 前置机的数据以 utf-8 编码；
e) 非结构化数据文件名中不允许出现“,”“&”“*”“/”等特殊字符。

6.4 服务接口归集

6.4.1 归集步骤

对于数据提供方存在现有接口或者具有开发能力的情况下，大数据资源平台可针对其进行接口的定制化开发，通过接口的方式进行数据传输。

6.4.2 技术要求

服务接口归集相关技术要求如下：

a) 归集频率：实时、分钟、小时等；
b) 归集类型：全量采集、增量采集；
c) 支持接口类型：https。

6.5 消息队列归集

6.5.1 归集步骤

数据提供方通过目录链管理系统将所需归集的数据注册编目，并发起数据归集任务，获得消息主题（topic）。数据提供方调用数据推送接口推送数据，大数据资源平台转发数据推送至消息主题（topic），并写入市级数据湖。

6.5.2 技术要求

消息队列归集相关技术要求如下：

a) 归集频率：实时；
b) 支持接口类型：https。

6.5.3 操作要求

流数据归集过程中应符合下列要求：

a) 接口请求仅需传入一个消息主题（topic）名称，一个消息主题（topic）对应一个数据目录，数据会持续写入对应的消息主题（topic）并写入市级数据湖；
b) 接口请求参数 datas 类型是列表，可传入多条数据，批量采集；
c) 接口请求参数 datas 项中，字段若为时间戳类型，应传 16 位微秒时间戳；
d) 接口请求参数 datas 项中，其 key 与消息主题（topic）中字段一致。

7 平台级联

7.1 市区级联

区大数据资源分平台应与市大数据资源平台实现级联互通，实现数据目录、数据标签、算法、数据服务、安全协同，形成上下标准统一、广泛覆盖、集中可控的共享开放渠道。

市区级联类型分为数据库表级联、文件级联和服务接口级联，见表1。

表1 市区级联类型

序号	级联类型	适用场景
1	数据库表级联	级联共享区资源为数据库表类型
2	文件级联	级联共享区资源为文件类型
3	服务接口级联	级联共享区资源为数据接口类型

7.2 长三角级联

大数据资源平台按照DB 31/T XXXX《长三角数据共享交换平台 数据接入规范》相关要求，与长三角数据共享交换平台实现级联对接，支持以库表接入、文件接入或服务接入的方式实现数据共享。

7.3 国省级联

市级大数据资源平台应按照GB/T 21062相关要求，与国家数据共享交换平台实现级联对接，同步数据目录，支撑按需调用，并提供统一规范的数据共享交换服务。

市级大数据资源平台应对数据共享过程进行追溯管理，支持问题数据的溯源定位和及时处理。

8 数据共享

数据共享管理应遵守DB31/T 1241-XXXX中第7章的规定。

9 安全保障

9.1 基本要求

大数据资源平台共享交换应满足以下基本要求：

a) 建立数据质量评估、数据共享应用成效评估、数据安全风险评估等安全保障制度，落实安全管理责任；
b) 加强数据共享交换过程中的监督管理，定期组织数据共享交换安全检查；

c） 满足GB/T 22239-2019中三级网络安全等级保护要求和GB/T 39477-2020 中共享数据换安全要求；

d） 根据GB/T 43697的数据分类分级要求，对共享交换的数据实施安全保护。

9.2 节点管理

前置交换节点应按要求进行安全防护，要求包括但不限于：

a） 使用防火墙、入侵检测系统等网络安全设备，按照最小化原则设置网络安全访问策略；

b） 记录前置交换系统的网络访问日志，以便对安全事件进行溯源和分析；

c） 采用安全的操作系统，并安装补丁和更新程序；

d） 采用基于角色的访问控制、最小权限原则等权限管理措施，确保对系统的访问权限最小化；

e） 对交换节点服务器和操作终端进行安全运维管理，主要操作可审计、可回溯。

9.3 权限控制

大数据资源平台应以最小化授权原则对数据提供方、数据使用方等进行权限控制，要求包括但不限于：

a） 以授权方式共享的，应明确用户使用角色、数据使用环境、访问权限等。涉及专库的，可按需拥有建表、删除自建表的权限；授权有时效限制，权限过期应重新申请授权；

b） 以非授权方式共享的，仅允许用在审核通过的业务场景和调用应用，不应再代理任何未经授权的其他接口；

c） 以跨层级数据返回、数据落地方式共享的，应控制返回或下发的数据的使用范围。

9.4 过程管控

应对公共数据共享交换过程进行安全管控，要求包括但不限于：

a） 数据提供方：

1） 以授权方式共享的，明确用户使用角色，确保生产与开发环境分离，对数据访问行为进行管控和接口风险监测；

2） 以非授权方式共享的，对接口 IP 白名单、接口调用数、接口有效期和接口加密通道进行合规检查；

3） 以跨层级数据返回、数据落地方式共享的，涉及敏感数据的表应加密处理，通过加密传输通道下发。

b） 数据使用方：

1） 以授权方式共享的，明确用户使用角色，确保生产与开发环境分离；

2） 以非授权方式共享的，对接口 IP 白名单、接口调用数、接口有效期和接口加密通道进行合规检查；

3） 以跨层级数据返回、数据落地方式共享的，采取加密传输通道抽取数据。

c） 大数据资源平台管理部门：

1）定期检查和评估平台共享交换数据传输的安全性和可靠性；

2）对数据共享交换通道运行情况实时监测告警，并及时处置；

3）实时监测服务接口调用，对调用频次异常、调用超时、调用错误、非授权调用等情况及时告警处置；

4）采用隐私计算、数据沙箱、数据水印等技术，加强数据使用安全管控；

5）对数据共享交换日志进行审计和分析。

参 考 文 献

[1] DB31/T 1446—2023 公共数据安全分级指南
[2] DB31DSJ/Z 004—2022 区级大数据资源平台建设指南
[3] 政务数据共享条例（中华人民共和国国务院令 第809号）
[4] 上海市数据条例（2022年1月1日）
[5] 全国一体化政务大数据体系建设指南（国办函〔2022〕102号）
[6] 上海市公共数据共享实施办法（试行）（2023年3月11日）
[7] 上海市公共数据目录链管理办法（沪数据办〔2025〕2号）