
上海教育信息化规范

编号：SJX-DAT/T 004-2021

上海教育市级数据资源管理技术平台 数据集成技术规范 (试行)

(本发布版最后修订日期： 2021-8-10)

上海市教育委员会信息中心

目 录

1.范围.....	1
2.参考资料.....	1
3.术语和定义.....	1
4.数据集成范围和内容.....	2
5.职责分工.....	2
6.数据集成技术规范.....	2
7.数据集成安全规范.....	7
附件一：接口调用示例.....	8
附件二：接口返回数据示例.....	11
附件三：部门（单位）代码命名参考规范.....	13
附件四：数据集成前置交换节点、前置交换库相关参数表.....	16

1. 范围

本规范规定了上海市教育数据集成的范围、内容、组织架构、职责及技术规范。

2. 参考资料

上海教育数据集成技术规范制定参考以下相关资料：

《上海教育数据管理办法（试行）》

3. 术语和定义

下列术语和定义适用于本文件。

3.1

数据交换

实现不同地域、不同计算机、不同软件的用户，获取、交换非本机数据并进行各种操作、运算和分析的过程。

3.2

数据共享

各部门因履行职责需要，使用其他部门的数据以及为其他部门提供数据的行为。

3.3

数据交换软件

用于数据、文件交换共享的软件。

3.4

前置交换节点

安装了数据交换软件、用于数据集成的前置机。

3.5

前置交换库

用于数据集成的数据库。

3.6

前置文件交换节点

用于文件集成的前置机。

4. 数据集成范围和内容

数据集成范围包括上海市教育委员会（以下简称“市教委”）、各区教育局及其辖区内的公办中小学和幼儿园、各市属公办高等学校（以下简称“各高等学校”） and 各市属公办中等职业学校（以下简称“各中职校”）等教育单位（部门）（以下简称“各级各类教育单位”）在履行职责过程中产生、采集和使用的各类非涉密数据。涉密数据管理按照国家和本市有关法律、法规进行。上海教育市级平台相关数据归集至教育部、大数据中心等单位时，遵照教育部及上海市相关数据地方标准执行。

5. 职责分工

根据《上海教育数据管理办法（试行）》，数据集成管理组织架构包括数据提供部门、数据使用部门、数据管理协调部门和数据技术管理部门。职责和权利由《上海教育数据管理办法（试行）》确定。

数据提供部门根据职能采集或产生某类数据，是该类数据的唯一权威来源，对该类数据有管理权和审核权。

数据使用部门因履行职责需要，可申请使用相关数据。

数据管理协调部门对数据集成、共享、服务和使用进行管理协调和监督审核。

数据技术管理部门为数据集成、共享、服务、使用及其管理过程提供技术支撑保障。

6. 数据集成技术规范

6.1 数据集成方式

数据集成的形式主要分为数据库表、文件和接口三类。三类形式的集成技术规范分别由 6.2、6.3 和 6.4 进行规定。

6.2 数据库表集成规范

数据库表集成是指以数据库表的形式进行的数据集成。数据提供部门需准备一个前置交换节点，用于安装数据交换软件；一个前置交换库，用于存储待归集的数据库表，并与前置交换节点连接。原则上一个单位只保留一个数据对接通道。

前置交换节点、前置交换库相关参数详见附件四。

6.2.1 数据库表命名规范

前置交换库中的数据库表是库表集成的基本元素，其命名应遵循以下规范：

- 1) 一般表采用“单位简称_系统名_表义名”格式构成；

如：jw_jjc_bmxt_xsjsx：单位简称（市教委基教处 jw_jjc）_系统名（报名系统 bmxt）_表义名（学生基本信息 xsjsx）；

- 2) 字典表采用“dict_单位简称_系统名_详细字典名称”格式构成；

如：dict_jw_jjc_bmxt_bmfs：dict_单位简称（市教委基教处）_系统名（报名系统 bmxt）_bmfs(报名方式)；

- 3) 临时表采用“单位简称_系统名_tmp_表义名”格式构成；

如：jw_jjc_bmxt_tmp_xsjsx：单位简称(市教委基教处)_系统名（报名系统 bmxt）_tmp_表义名（学生基本信息 xsjsx）；

- 4) 命名不得超过 30 个字符；对象名的字符之间不得含有空格；命名不得与保留词、数据库系统或常用访问方法冲突。

6.2.2 数据库表建表规范

由于各数据提供部门的原始库类型不同、数据类型不同，应对前置交换库中的表结构进行规范性定义。规范性定义应在调研了解原始系统库表结构的基础上，结合国家相关规范和数据共享需要，采用数据映射和转换等方法，对原始的字段名、字段类型、字段格式、值域等重新定义，要求如下：

- 1) 每张表单必须有主键，主键不可与业务相关，主键类型为文本类型，长度不超过 32 位。
- 2) 每张表单要定义时间戳字段“generatorTime”，用于记录业务更新时间，即数据在原系统的生成时间，时间类型为 timestamp 类型，精确到毫秒级。
- 3) 每张表单要定义时间戳字段“createTime”，用于记录数据创建时间，即数据完成归集时间，时间类型为 timestamp 类型，精确到毫秒级。
- 4) 每张表单要定义时间戳字段“updateTime”，用于记录最后更新时间，时间类型为 timestamp 类型，精确到毫秒级。
- 5) 每张表单要定义标识字段“flag”，用于标识数据状态，1 表示新增（初次插入数据），2 表

示更新（数据有变更），3 表示删除（数据被删除）。

6.2.3 数据库表读写规范

对于数据库表，前置交换节点通过数据库适配器以 JDBC 的方式接入前置交换库，对库中的数据进行读写。

数据库表读取，要求前置交换节点具有对该表的查询权限。除特殊要求，不授予写入和更新的权限。

数据库表写入，要求前置交换节点具有对该表的读取、写入、更新权限。

数据库表写入流程主要分为全量写入和增量写入两类。全量写入流程仅用于初次写入的数据库表，对数据库表记录进行全量读取；增量写入用于获取日常产生的增量数据，根据标志位及时间戳来判断增量数据，最大程度减少对数据库性能的影响。

数据写入频度默认按天进行读写，每天读写一次数据，最高支持按分钟进行读写。

对于前置交换库写入过程要遵循以下规范：

- 1) 对于初次写入，将待集成的数据全量同步至前置交换库中，默认同步标识 flag=1（新增），并将同步时间写入时间戳字段（updateTime）。
- 2) 对于增量数据的写入，应通过主键判断数据行是否存在，如不存在，可直接写入一条新纪录，默认标识 flag=1（新增），如果数据已存在，判断数据是否有变化。如无变化，标志位不变；如有变化，更新 flag=2（更新）；如需删除数据，更新 flag=3（删除），并将同步时间写入时间戳字段（updateTime）。
- 3) 禁止对前置交换库中的数据进行物理删除。
- 4) 对于已有数据的表单，禁止使用全表删除再全表写入的方式进行数据更新。
- 5) 视图作为一种特殊的虚拟表，也属于库表读取的一种形式。可建立查询视图，以只读视图的形式提供复杂多表关联查询的数据结果。视图的查询结果应包含时间戳字段，以便于定期获取增量数据。

6.3 文件集成规范

文件集成是指将文件上传至前置文件交换节点指定的文件目录，由前置文件交换节点进行归集的数据集成。文件主要分为格式化文件（Excel、TXT、CSV、XML 等）与非格式化文件（照片、Word 等）。

6.3.1 目录命名规范

数据提供部门待归集的文件数据存放在前置文件交换节点的文件存储区中指定文件目录。文件目录应按照“数据提供部门_系统名”_文件集名”规则命名，同类文件都应存放在同一目录下（必要时可设置子目录）。例如，来自市教委基教处所属报名系统的学生照片，目录名应为“jw_jjc_bmxt_stuphotos”。

每个文件目录对应设置备份目录，用于存放已归集过的文件备查，目录按照“数据提供部门_系统名_文件集名_bak”规则命名，如 jw_jjc_bmxt_stuphotos_bak。

部门（单位）代码命名参考规范详见附件三。

6.3.2 非格式化文件集成规范

对于非格式化文件，由数据提供部门定期上传到前置文件交换节点中指定的文件目录分类存放。前置交换节点通过定期扫描相应文件夹，将文件通过交换软件(SFTP 协议)传输至市级教育数据资源管理技术平台。

前置交换节点归集成功的文件，应在原目录删除并备份至相应的 bak 目录备查，备份时应在原文件名后添加时间戳，防止文件重名。

6.3.3 格式化文件集成规范

对于格式化文件（Excel，txt，CSV 等），应提前约定文件类型，宜以 Excel 及 CSV 文件为主。对于文件的内容，应事先制定文件格式模板，包含字段名、字段顺序、字段类型、字段长度等数据元信息。数据提供部门在提供文件时，必须严格按照文件模板格式生成文件。

前置交换节点通过定期扫描相应文件夹，读取并解析格式化文件内容，转化为数据库可识别的格式，通过 Http 协议传递给市级教育数据资源管理技术平台。

前置交换节点归集成功的文件，应在原目录删除并备份至相应的 bak 目录备查，备份时应在原文件名后添加时间戳，防止文件重名。

6.4 接口集成规范

接口集成指通过提供开放接口的形式进行数据集成。

6.4.1 接口命名规范

接口按照“数据提供部门_具体业务”规则命名，如国际交流处的中外合作办学名单，接口命名为 jw_gjjlc_zwhzbx。

方法命名要根据操作类型进行区分，如：获取：get()；更新：update()；新增：create()；删

除：delete()；统计：count()；保存：save()。如通过接口获取国际交流处的中外合作办学名单：

https://apigw.shec.edu.c/shec/jw_gjjlc_zwhzbx/getZwhzbxmd。

6.4.2 接口入参规范

接口应符合 Soap1.2 协议的 Webservice 接口。接口头信息需包含如下参数：

参数名	说明	备注
OrgName	接口提供单位名称	教委统一规范
OrgCode	接口提供单位代码	教委统一规范
BusinessName	业务系统名称	发送方指定
BusinessCode	业务系统编码	发送方指定
Query	查询条件	根据具体的数据交换业务定制化
TableName	库表名	需要交换的库表名称，发送方指定

接口调用示例详见附件一。

6.4.3 接口返回规范

接口返回值标识的规范定义如下：

标签	定义
<DataSet></DataSet>	结果集标签
<RecordData></RecordData>	数据集标签
<UnitData></UnitData>	字段标签
<UnitIDName></UnitIDName>	字段名称标签
<UnitDisplayName></UnitDisplayName>	字段中文名标签
<UnitValue></UnitValue>	字段值标签

接口返回值样例详见附件二。

6.4.4 接口认证规范

为保障各数据提供部门开发的接口被安全调用，接口的头信息部分应添加用户名及密码的验证，各部门在做接口接入申请时应告知正确的用户名、密码。

接口调用设置的密码应符合一定的复杂性要求，必须使用字母、数字及特殊字符组合的方式，且长度不少于 20 位。

各部门开放的接口，应固定端口号（禁止使用常用的 80、8080 端口），同时接口仅允许来自于该部门前置交换节点所在 IP 发出的调用请求。

前置交换节点在调用接口时，将获知的用户名、密码内容通过 MD5+base64 加密后传入头信息，校验通过后，方可实现成功调用。

7. 数据集成安全规范

数据集成过程安全关注数据从数据源到数据宿过程的保密性、完整性、可用性。涉及到国家重要信息、单位敏感信息和个人隐私信息的数据传输场景，应进行加密传输。在网络传输过程中，防止数据通过各种网络应用泄露。

具体参见《上海教育数据安全规范（试行）》。

附件一：接口调用示例

```
package dataIntegration;

import java.io.UnsupportedEncodingException;
import java.util.Base64;
import org.apache.commons.codec.digest.DigestUtils;
import org.codehaus.xfire.MessageContext;
import org.codehaus.xfire.handler.AbstractHandler;
import org.jdom.Element;

public class ClientAuthenticationHandler extends AbstractHandler {
    private String username = null;
    private String password = null;
    public ClientAuthenticationHandler() {
    }
    public ClientAuthenticationHandler(String username, String password) {
        this.username = username;
        this.password = password;
    }
    public void setUsername(String username) {
        this.username = username;
    }
    public void setPassword(String password) {
        this.password = password;
    }
    public void invoke(MessageContext context) throws Exception {
        Element el = new Element("header");
        context.getOutMessage().setHeader(el);
        Element auth = new Element("AuthenticationToken");
        Element username_el = new Element("Username");
        username_el.addContent(encryption(username));
        Element password_el = new Element("Password");
        password_el.addContent(encryption(password));
        auth.addContent(username_el);
        auth.addContent(password_el);
        el.addContent(auth);
    }
    //通过MD5+base64加密
    private String encryption(String str) {
        String base64encodedString = "";
        try {
            base64encodedString =
                Base64.getUrlEncoder().encodeToString(DigestUtils.md5Hex(str).getBytes("utf-8"));
        }
    }
}
```

```

        catch (UnsupportedEncodingException e) {
            e.printStackTrace();
        }
        return base64encodedString;
    }
}

package dataIntegration;
import java.lang.reflect.Proxy;
import java.util.HashMap;
import java.util.Map;

import org.codehaus.xfire.client.Client;
import org.codehaus.xfire.client.XFireProxy;
import org.codehaus.xfire.client.XFireProxyFactory;
import org.codehaus.xfire.service.Service;
import org.codehaus.xfire.service.binding.ObjectServiceFactory;
import org.json.JSONArray;
public class DataIntegrationTest {
    /**
     * @param args
     */
    public static void main(String[] args) {
        try{
            Service serviceModel = new
ObjectServiceFactory().create(DataIntegrationWebService.class);
            DataIntegrationWebService service = (DataIntegrationWebService) new
XFireProxyFactory().create(serviceModel,
                "http://localhost/webservice/DataIntegrationWebService");
            XFireProxy proxy = (XFireProxy)Proxy.getInvocationHandler(service);
            Client client = proxy.getClient();
            //发送授权信息
            client.addOutHandler(new ClientAuthenticationHandler("用户名","密码"));
            //添加信息
            Map<String, Object> addMap = new HashMap<String, Object>();
            addMap.put("OrgName", "OrgName");//接口提供单位名称-教委统一规范
            addMap.put("OrgCode", "OrgCode");//接口提供单位代码-教委统一规范
            addMap.put("BusinessName", "BusinessName");//业务系统名称-发送方指定
            addMap.put("BusinessCode", "BusinessCode");//业务系统编码-发送方指定
            addMap.put("Query", "Query");//查询条件-根据具体的数据交换业务定制化
            addMap.put("TableName", "Query");//库表名-需要交换的库表名称, 发送方指定
            JSONArray addJson = new JSONArray();
            addJson.put(addMap);

```

```

        Object[] addResult = client.invoke("getInfo", new Object[] {addJson});
        String xmlResult = addResult[0].toString();
        System.out.println(xmlResult);
    } catch (Exception e) {
        e.printStackTrace();
    }
}

package dataIntegration;

public interface DataIntegrationWebService {
    public String getInfo(String paramjson);
}

```

附件二：接口返回数据示例

```
<?xml version="1.0" encoding="GB18030"?>
<BusinessData xmlns="http://www.egs.org.cn/businessdata"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.egs.org.cn/businessdata.xsd">
  <DataSet>
    <RecordData>
      <UnitData>
        <UnitIDName>ID</UnitIDName>
        <UnitDisplayName>学生身份号码</UnitDisplayName>
        <UnitValue>110110199503231463</UnitValue>
      </UnitData>
      <UnitData>
        <UnitIDName>Name</UnitIDName>
        <UnitDisplayName>姓名</UnitDisplayName>
        <UnitValue>张三</UnitValue>
      </UnitData>
      <UnitData>
        <UnitIDName>Sex</UnitIDName>
        <UnitDisplayName>性别</UnitDisplayName>
        <UnitValue>男</UnitValue>
      </UnitData>
      <UnitData>
        <UnitIDName>Ethnic</UnitIDName>
        <UnitDisplayName>民族</UnitDisplayName>
        <UnitValue>汉</UnitValue>
      </UnitData>
      <UnitData>
        <UnitIDName>Birthday</UnitIDName>
        <UnitDisplayName>出生日期</UnitDisplayName>
        <UnitValue>19950323</UnitValue>
      </UnitData>
      <UnitData>
        <UnitIDName>Address</UnitIDName>
        <UnitDisplayName>常住户口所在地住址</UnitDisplayName>
        <UnitValue>上海市长宁区xx小区xx号</UnitValue>
      </UnitData>
    </RecordData>
    <RecordData>
      <UnitData>
        <UnitIDName>ID</UnitIDName>
        <UnitDisplayName>学生身份号码</UnitDisplayName>
```

```
<UnitValue>110110199503231464</UnitValue>
</UnitData>
<UnitData>
<UnitIDName>Name</UnitIDName>
<UnitDisplayName>姓名</UnitDisplayName>
<UnitValue>李四</UnitValue>
</UnitData>
<UnitData>
<UnitIDName>Sex</UnitIDName>
<UnitDisplayName>性别</UnitDisplayName>
<UnitValue>女</UnitValue>
</UnitData>
<UnitData>
<UnitIDName>Ethnic</UnitIDName>
<UnitDisplayName>民族</UnitDisplayName>
<UnitValue>汉</UnitValue>
</UnitData>
<UnitData>
<UnitIDName>Birthday</UnitIDName>
<UnitDisplayName>出生日期</UnitDisplayName>
<UnitValue>19950323</UnitValue>
</UnitData>
<UnitData>
<UnitIDName>Address</UnitIDName>
<UnitDisplayName>常住户口所在地住址</UnitDisplayName>
<UnitValue>上海市长宁区xx小区xx号</UnitValue>
</UnitData>
</RecordData>
</DataSet>
</BusinessData>
```

附件三：部门（单位）代码命名参考规范

为统一“数据存储目录名称”定义规则中“文件名称”的代码表示，给予部门（单位）代码命名的参考规则，机构汉字简称的首字母建议不超过 5 位：

部门（单位）类别	代码规则	示例
市教委内设机构	jw_机构汉字简称的首字母	办公室：jw_bgs 基教处：jw_jjc
市教委机关及直属事业单位	jw_机构汉字简称的首字母	上海老年大学：jw_lndx 上海市教育人才交流服务中心：jw_rczx
区教育局内设机构	jw_区名首字母_机构汉字简称的首字母	黄浦区教育局组织科：jw_hp_zzk 黄浦区教育局纪检监察科：jw_hp_jjk
区教育局机关及直属事业单位	jw_区名首字母_机构汉字简称的首字母	黄浦区业余大学： jw_hp_yydx 黄浦区教育人才服务中心： jw_hp_jrfz
公办中小学	jw_所在区名首字母_zxx_机构汉字简称的首字母	格致中学：jw_hp_zxx_gzzx 上海市黄浦区第一中心小学： jw_hp_zxx_dyzzxx
公办幼儿园	jw_所在区名首字母_kin_机构汉字简称的首字母	上海市黄浦区星光幼儿园： jw_hp_kin_xgyey
市属公办高等学校	jw_u_机构汉字简称的首字母	上海科技大学： jw_u_skd
公办中等职业学校	jw_zzpt	jw_zzpt（即中职平台）

各部门（单位）名称若发生变更，需及时执行变更流程。

上海市教育委员会及其部分直属事业单位的代码可参照下表：

单位、机构所属类别 名称	单位、机构名称	单位、机构代码
上海市教育委员会内 设机构	办公室	jw_bgs
	秘书处（研究室）	jw_msc
	发展规划处	jw_ghc
	人事处	jw_rsc
	财务处（国有资产管理处）	jw_cwc
	基础教育处	jw_jjc
	职业教育处	jw_zyjc
	高等教育处	jw_gjc
	终身教育处	jw_zsjc
	托幼工作处	jw_tyc
	民办教育管理处	jw_mbc
	学生处	jw_xsc
	科学技术处	jw_kjc
	信息化工作处	jw_xxhc
	体育卫生艺术科普处	jw_twykc
	国际交流处（港澳台办公室）	jw_gjjlc
	学校后勤保卫处	jw_hbc
	政策法规处（政风行风办公室）	jw_fgc
	教材和语言文字管理处	jw_ygc
	审计处	jw_sjc
	德育处	jw_dyc
	青少年保护工作处	jw_qbc
	信访办公室	jw_xfb
上海市教育委员会直 属事业单位	上海教育报刊总社	jw_bkzs
	上海市教委教研室	jw_jys
	上海市教育评估院	jw_pgy
	上海市学生事务中心	jw_xszx
	上海市教育人才交流服务中心	jw_rczx
	上海市教委科技发展中心科发中心	jw_kfzx
	上海市教委财务与资产管理中心	jw_czzx
	上海市教育基建管理中心	jw_jjzx
	上海市教委信息中心	jw_xxzx
	上海市科技艺术教育中心	jw_kyzz
	上海市教委教育技术装备中心	jw_zbzx
	上海市青少年校外活动营地	jw_dflz
	上海市语言文字测试中心	jw_yczx
	上海市教育督导事务中心	jw_ddzx

	上海高校后勤服务中心	jw_hqzx
	上海教育会堂	jw_jyht
	上海市师资培训中心	jw_spzx
	上海老年大学	jw_lndx
	上海中学	jw_shzx
	上海市行政管理学校	jw_xzgl
	上海市聋哑青年技术学校	jw_lqjx
	上海市盲童学校	jw_mtxx

附件四：数据集成前置交换节点、前置交换库相关参数表

基本信息：

No.	项目	内容	备注
1	前置机归属单位		
2	相关业务系统 / 数据中心		
3	相关负责人 / 联系方式		
4	相关技术对接人 / 联系方式		

前置交换机（虚机）：

No.	项目	内容	备注
1	操作系统	Linux 操作系统，建议为 CentOS 7.x 或更高版本	
2	配置要求	4 Core CPU、16GB 内存、200GB 磁盘	
3	IP：SSH 管理端口	请根据实际服务器情况填写	供信息中心远程安装交换软件使用
4	管理员账号	请根据实际服务器情况填写	
5	管理员密码	请根据实际服务器情况填写	

前置交换库（RDS）：

No.	项目	内容	备注
1	数据库	建议选择 mysql5.6 或更高版本、oracle11g 或更高版本、sqlserver2008r2 或更高版本、达梦数据库最新版本。	
2	IP：数据库端口	请根据实际数据库情况填写	供信息中心配置交换软件访问数据库中使用
3	Database（Schema）	请根据实际数据库情况填写	
4	用户名	请根据实际数据库情况填写	
5	密码	请根据实际数据库情况填写	

网络安全配置：

1、应配置网络防火墙：允许市级教育数据资源管理技术平台访问前置交换节点；允许前置交换节点访问市级教育数据资源管理技术平台；允许市级远程管理机访问前置交换节点，保证服务器的入口外网 IP 与出口 IP 保持一致，针对市教委服务器进行白名单设置。

A. 如前置交换节点位于上海教育数据中心：

源 IP	目的 IP、端口	策略
市级教育数据资源管理技术平台内网地址： 172.31.24.1、172.31.24.2	前置交换节点内网 IP 地址，SSH 访问端口、端口 8080	允许
市级远程管理机的内网 IP 地址：172.31.25.3	前置交换节点内网 IP 地址，SSH 访问端口、8080 端口（web 访问）	允许
前置交换节点的内网 IP 地址	172.31.24.1 的 8080 端口，以及 172.31.24.2 的 8080 端口	允许

B. 如前置交换节点不位于上海教育数据中心，但有校校通网络出口：

源 IP	目的 IP、端口	策略
市级教育数据资源管理技术平台校校通地址： 10.0.255.74	前置交换节点校校通IP地址，SSH访问端口、 8080端口	允许
市级远程管理机的校校通地址：10.0.255.90	前置交换节点的校校通IP地址，SSH访问端口、 8080端口（web访问）	允许
前置交换节点的校校通 IP 地址	10.0.255.74的8080端口以及8082端口	允许

C. 如前置交换节点不位于上海教育数据中心，且仅有互联网出口：

源 IP	目的 IP、端口	策略
市级教育数据资源管理技术平台城域网（互联网）地址：202.120.199.97	前置交换节点互联网IP地址，SSH访问端口、 8080端口	允许
市级远程管理机的城域网（互联网）地址： 202.120.199.9	前置交换节点的互联网IP地址，SSH访问端口、 8080端口（web访问）	允许
前置交换节点的互联网 IP 地址	202.120.199.97的8080以及8082端口	允许

2、前置交换库应允许前置交换节点访问（前置交换库不需要对外）

源 IP（前置交换节点）	目的 IP、端口（前置交换库）	策略
前置交换节点内网 IP 地址	前置交换库内网IP地址，数据库端口	允许