



四川大學
SICHUAN UNIVERSITY

高校数据治理的 挑战、实践与思考

四川大学信息化建设与管理办公室 段磊

2023.11 福州



- 国际数据管理协会（The Data Management Association）：数据治理是**对数据资产管理行使权力和控制的活动集合**。
- 百度：数据治理是组织中**涉及数据使用的一整套管理行为**。
 - 由企业数据治理部门发起并推行，关于如何制定和实施针对整个企业内部数据的商业应用和技术管理的一系列政策和流程。
- 国际数据治理研究所（Data Governance Institute）：数据治理是一个**通过一系列信息相关的过程来实现决策权和职责分工**的系统。
 - 这些过程按照达成共识的模型来执行，该模型描述了谁（**Who**）能根据什么信息，在什么时间（**When**）和情况（**Where**）下，用什么方法（**How**），采取什么行动（**What**）。

目标

持续提升学校数据资产的质量并挖掘其价值

范围

机构与人员、流程与制度、执行与管控

途径

适合校情的**数据管理体系**的搭建与持续优化



1 数据可得

找到从数据中挖掘价值的明确途径

数据采集

2 数据可靠

基于数据确权和数据定价进行流通交易

数据流通

3 数据可用

发挥数据的要素价值，产生更大的生产力

数据要素

高校“数据孤岛”出现的必然性



教务系统



财务系统



学工系统



科研系统



办公系统

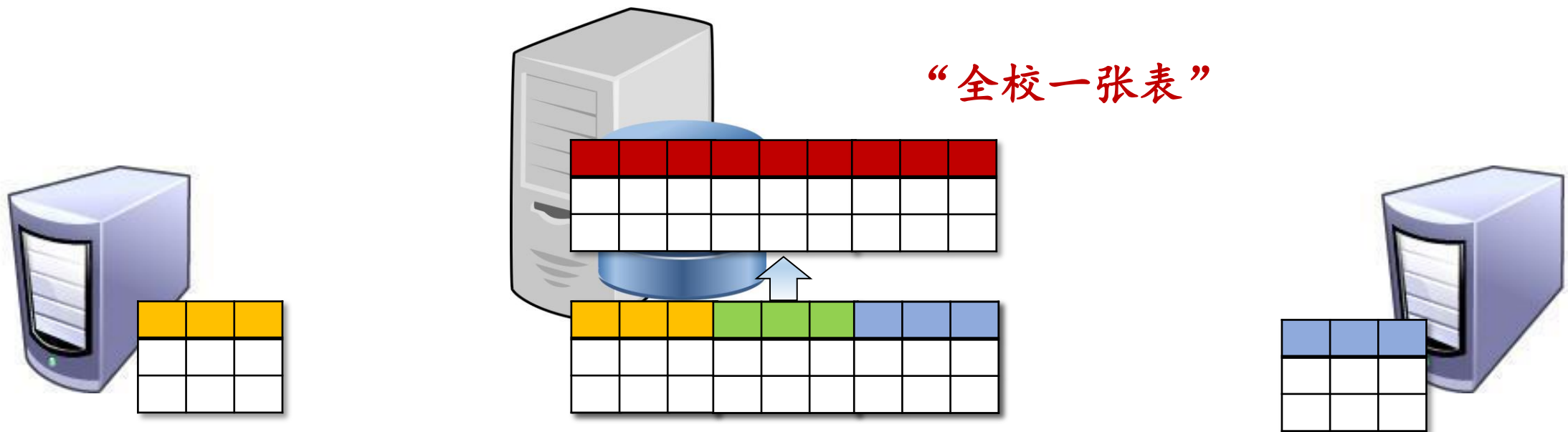


一卡通系统

...

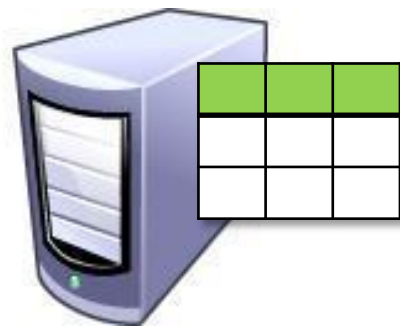
管理信息系统 (Management Information System, MIS) 是一个以人为主导, 利用计算机硬件、软件、网络通信设备以及其他办公设备, 进行信息的收集、传输、加工、储存、更新、拓展和维护的系统。

基于MIS系统的校园信息化必然出现“数据孤岛”



建设数据中台

- 一数一源路径清晰
- 支持多源数据融合
- 统一数据管理



功能模块

非实时场景的批式
数据交换与集成



离线开发

一站式多引擎离线数据开发平台，新一代数据仓库，开发效率升一倍

实时场景的流式
数据交换与集成



实时开发

全链路流计算FlinkSQL开发平台，秒级实时开发，人力省一半



算法开发

可视化的机器学习算法开发平台，算法开发工程化，效率提升 100%



数据服务

可视化生成、注册、管理数据API，快速构建 OneService 数据共享服务

快速构建多样的
数据共享服务

数据挖掘算法开发
运行平台

元数据、数据标准、
数据模型管理



数据资产

实现企业数据的标准化与资产化，打造企业级数据中心

数据一致性和完整性
校验规则管理



数据质量

配置关键数据表校验与监控规则，及时发现数据质量问题



智能标签

基于 OneID 构建的实体标签引擎，数据萃取赋能业务创新



控制台

多集群管理与计算资源分配，帮助企业构建安全稳定的集群环境

数据中台集群管理
与资源分配

构建实体标签模型
抽象业务过程

川大数据中台：基于大数据实践的数据管理平台型工具，支持PB级的数据计算、存储与分析。

系统展示——川大数据中台

系统管理 数据源中心

BatchWorks 四川大学... 数据源 数据开发 运维中心 数据地图 项目管理

EasyAPI 川大API 概览 API市场 API管理 数据源管理 项目管理

BatchWorks 项目列表 数据地图

DataAssets 资产盘点 元数据 数据资源 数据标准 数据模型 平台管理

BatchWorks 四川大... 数据源 数据开发 运维中心 数据地图 项目管理

AIWorks 算法测试 数据管理 算法开发 运维中心 模型中心 项目管理

数据资产概况 (当前租户内的数据,截至2021-12-18) 只看我参与的项目

全局概览 我的概览

累计调用 795 次

失败率 0 %

接口总数 ---

错误类型分布

数据地图分布

分布属性: 负责人

今日周期实例执行汇总

任务失败 2

今日周期实例完成情况

数据挖掘

数据价值排行

排序	表名
1	ods_rsxt_api...
2	dm_zhkt_kh...

执行时长排行

任务名称
api2hive_ods_rsxt_api_user_depar...
dm_tgsq_lykt_info_df
oracle_2hive_ODS_TSGLXT_v_loanh...

cluster_thermal

读取数据表_1

类型转化_1

oneHot编码_1

pca主成分分析_1

Python脚本_1

kmeans聚类_1

预测_1

2023/11/30

提供数据服务

- 师生位置计算
- 校园消费分析



数据计算



数据融合

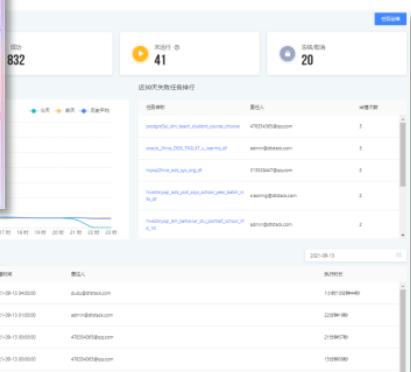


- 高基表填报
- 新生大数据

数据同步



- 数据共享
- 实时统计



数据交换



- 身份认证
- 门禁管理

整理出覆盖全校业务的八大领域信息标准规范，是数据环境稳定性、独立性的基础，
是信息系统实现信息共享集成和最优化集成的根本保证。



数据资产门户

四川大学数据资产目录

SICHUAN UNIVERSITY DATA ASSET CATALOGUE

请输入搜索内容

全站搜索



41

部门



34

系统



10

数据域



312

数据表



5174

字段数量

领域分类

部门分类



学生管理



教学管理



教职工管理



科研管理



财务管理



外事管理



ASSETS



党建引领



办学资源



人才培养



师资队伍



学科建设



科学研究



国际交流



社会服务



思想教育

领域分类

主题分类

部门分类

主题分类

部门分类



党委统战部



校招生办公室



党委学生工作部



机关党委



人事处



发展规划处



研究生院



实验室及设备管理处



国资管理处



老干部党总支、离退休工作处



基建处



信息化建设与管理办公室



后勤保障部



对外联络办公室



心理健康教育中心



档案管理



图书馆



社区建设办公室



分析测试中心



出版社



期刊社



博物馆



实验动物中心



成人继续教育学院



国家留学人员培训部 (出国...)



文化科技协同创新研发中心



国家生物医学材料工程 技术...



出国留学人员培训部



人才交流中心代管人员



就业指导中心

坚持一数一源一责任单位

- 一类数据原则上只应有一个数据源
- 通过梳理资源目录明确数据源责任单位
- 最大化减少数据重复采集
- 鼓励数据共享，提升数据应用的效能
- 提供数据接口和开放数据目录，依规申请使用

数据使用坚持最小必要

- 数据收集遵循最小够用原则
- 数据查看遵循最小授权原则
- 数据存储遵循最短周期原则
- 数据共享遵循用而不存原则



01

校内部处合作的数据治理

- 与图书馆进行论文类数据治理实践

02

与教育部合作的数据治理

- 与教育部教育管理信息中心共同搭建“高校数据共享应用平台”

03

校企间合作的数据治理

- 与天府通定制教职工往返江安校区地铁出行方案

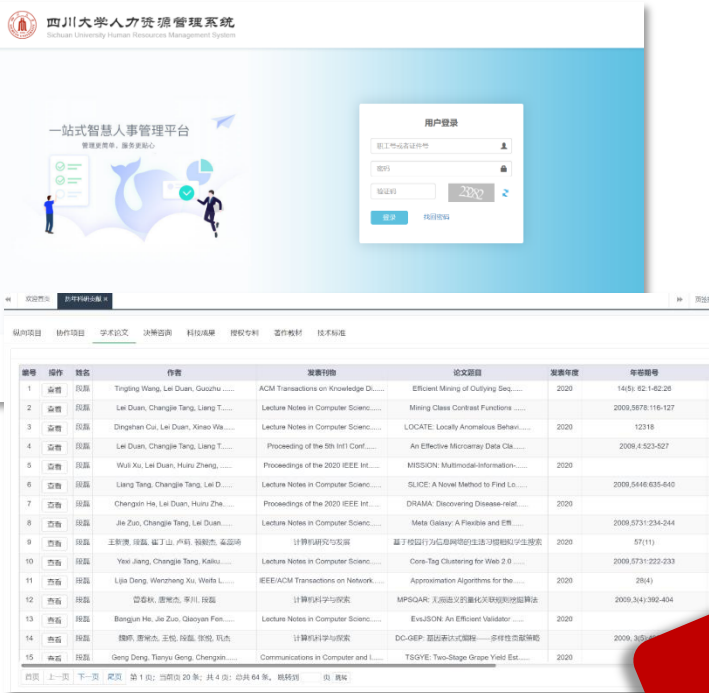
论文、专利、专著类科研成果数据是重要的校情数据，对
学科发展分析、教师考核评聘、学生培养评价等有着重要作用

两个关键问题：

如何获取？ —— **数据来源**

如何流通？ —— **数据共享**

以论文类数据治理为例



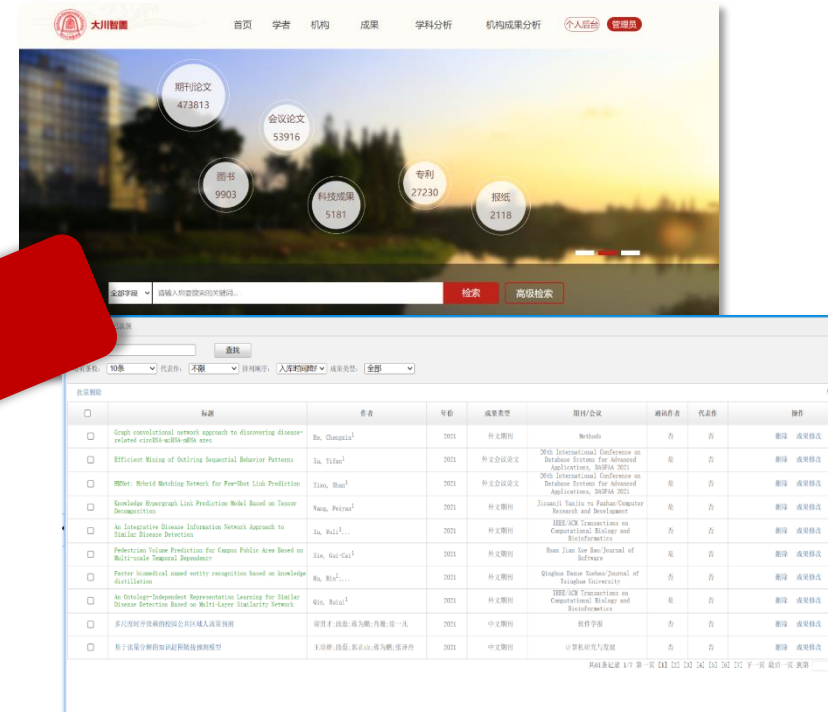
人力资源管理系统

- 面向全校
- 考核期间填报
- 无分选功能
- 手动填报



科研管理信息系统

- 面向理工医
- 常态化填报
- 部分分选功能
- 手动填报



机构与学者知识库

- 面向全校
- 常态化填报
- 分选功能
- 手动填报



- 20,000+校内**学者**
- 570,000条科研**成果**记录
- 92个校内**二级机构**
- 期刊、会议论文、专利、图书、报纸
- 所有论文成果均标注**科研评价**（期刊级别）
- 150,000条科研成果已**自动核定**
- 支持学校排行、学科排行、学科热点、国际竞争力分析

数据统计截至2022年3月

“四川大学机构与学者知识库”是**图书馆**依据《高等学校数字校园建设规范》建设的学术成果（科研资源）仓储系统。

一数一源一责任单位：

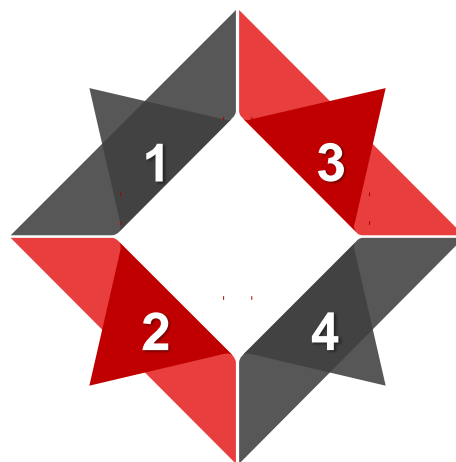
- 将“机构与学者知识库”作为论文类数据源头
- 将图书馆作为论文类数据的责任单位

数据基础好

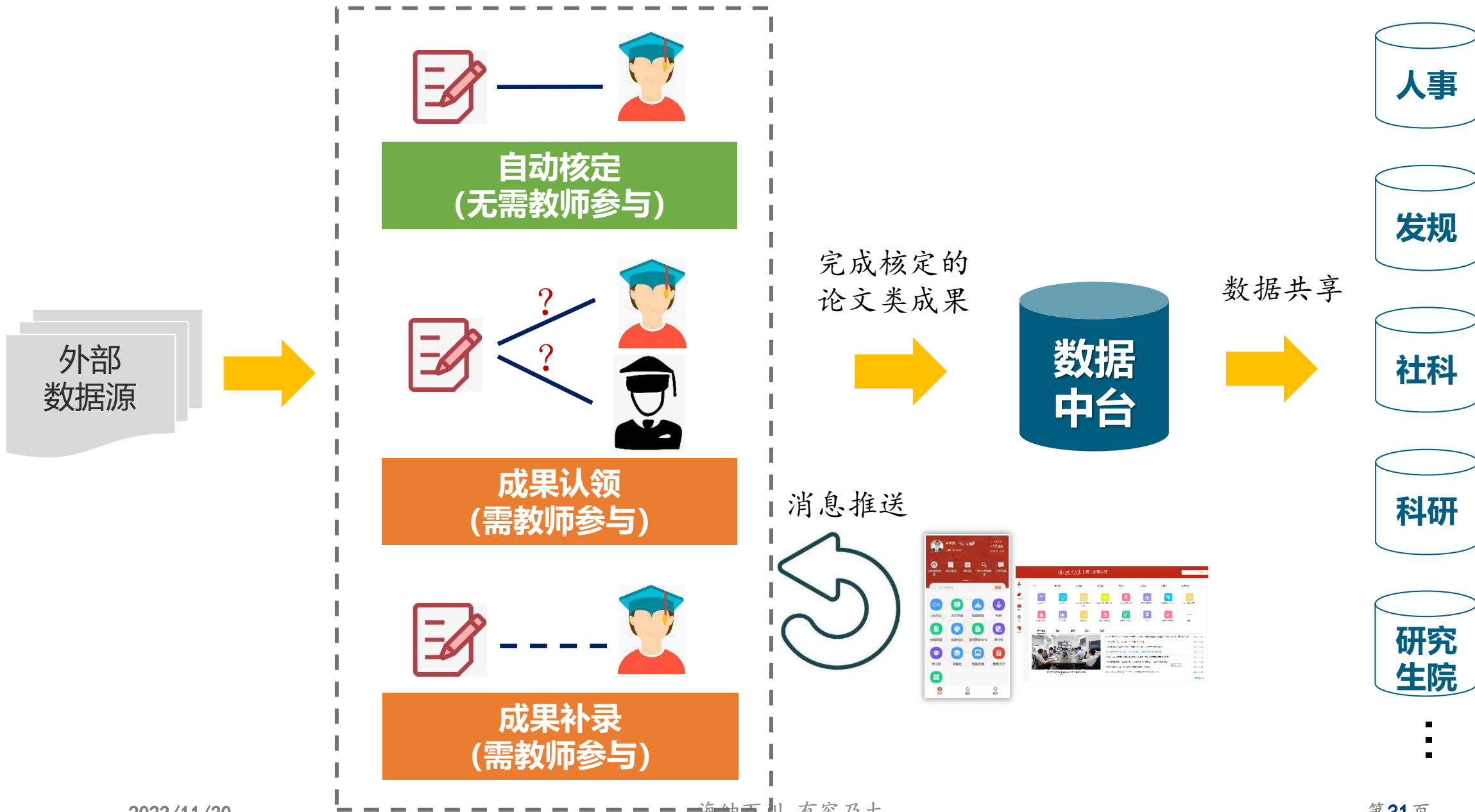
时效性强

准确度高

人力有保障



解决方案：最多填报一次



论文、专利、专著类科研成果数据是重要的校情数据，对
学科发展分析、教师考核评聘、学生培养评价等有着重要作用

两个关键问题：

如何获取？——数据来源

确定图书馆为论文类数据的采集源头，杜绝多头采集

如何流通？——数据共享

确定由数据中台提供论文类数据共享，避免重复填报

01

校内部处合作的数据治理

- 与图书馆进行论文类数据治理实践

02

与教育部合作的数据治理

- 与教育部教育管理信息中心共同搭建“高校数据共享应用平台”

03

校企间合作的数据治理

- 与天府通定制教职工往返江安校区地铁出行方案



2023年2月23日，四川大学作为首批24所高校之一，与教育部教育管理信息中心签署战略合作协议，共同搭建“高校数据共享应用平台”。该平台在全面整合教育部数据资源的基础上，汇聚包括公安部、民政部等20余个部委的100多个服务于高校业务的数据资源，可支持不同场景、不同形态的数据服务。



信息 核验

高校管理是学校建设中不可或缺的重要组成部分。为实现数字化的师生服务，**信息核验是重要环节之一**，应用范围广泛，包括以下几种实际场景：

- ✓ 新生入学时对身份信息、招生照片以及入校采集照片进行逐一比对，避免身份冒用等不良行为；
- ✓ 对师生在各业务系统上传或更换的个人照片进行核验，确保照片的有效性和真实性，杜绝师生身份冒用行为；
- ✓ 校园卡身份识别过程中教职工离职、去世等状态信息更新周期较长，在编人员或无固定期限人员等校园卡长期有效，存在非法人员使用去世教职工校园卡刷门禁实施盗窃、非法闯入等恶意行为的隐患。



数据 共享

目前师生身份信息的核验大多采取线下（人工）方式，不仅流程繁琐、更新不及时，耗费大量的时间和人力，且存在一定错误率。**四川大学通过数据共享实现核验流程再造，完成对师生身份信息和状态的实时校验，保证业务系统的及时更新。**

场景一 新生身份核验

1. 将新生身份信息从数据中台共享至人脸信息库。
2. 人脸信息库通过学生身份证号、学生姓名、学生照片与高校共享应用平台“CTID平台_身份认证请求接口”进行数据交互，实现对新生身份信息、入校采集照片和招生照片的比对，完成新生身份信息的核验。



场景二 师生证件照核验

1. 师生通过学校的移动端“云上川大”APP上传和更换证件照片。
2. “云上川大”APP通过API接口将采集的照片及身份信息与人脸信息库中信息进行比对，如人脸信息库中无此照片，再通过高校共享应用平台“CTID平台_身份认证请求接口”进行比对，从而实现师生证件照上传、更换的自动精准核验。



场景三 校园卡有效性核验

1. 数据中台通过对接人事管理系统和高校共享应用平台“公安部_人口库_身份核查”API接口对教职工在岗状态（在岗、不在岗）、人员类型（离校、离世）等信息进行周期更新。
2. 将更新的信息通过API接口共享至校园卡管理系统。校园卡管理系统及时更新校园卡状态，将离职或离世教职工的校园卡更新为停用状态，确保校园卡的合规使用。



01

校内部处合作的数据治理

- 与图书馆进行论文类数据治理实践

02

与教育部合作的数据治理

- 与教育部教育管理信息中心共同搭建“高校数据共享应用平台”

03

校企间合作的数据治理

- 与天府通定制教职工往返江安校区地铁出行方案



四川 大学

有望江、华西和江安三个校区，其中江安校区与望江校区相距约16公里。

2020年底，成都地铁8号线开通，设“川大望江校区站”和“川大江安校区站”，单程跨



通勤 需求

1. 如何对参与优惠活动的教职工进行身份识别？
2. 如何采集优惠活动所需要的教职工身份信息？
3. 如何安全地与天府通公司同步更新相关信息？
4. 如何从天府通公司获得教职工乘车明细数据？

通过信
职工乘坐



地铁 优惠

四川大学在职教职工，通过四川大学“云上川大”APP网上申请优惠并在“天府通”APP实名认证后，在“天府通”APP里获得**四川大学教职工专属身份标识**，使用“天府通”APP**电子钱包支付地铁乘车费用**时，享受四川大学教职工**优惠减扣**。

数据赋能——携手天府通定制教职工地铁出行



川大教职工通过“云上川大”APP【江安地铁乘车】发起申请



云上川大APP

用户姓名
手机号码

数字人民币专线

手机号码
乘车明细

向财务处提供乘车明细并协助财务处完成优惠扣费的补贴

- ✓ 完成实名认证
- ✓ 四川大学教职工标识
- ✓ 天府通乘车码
- ✓ 天府通电子钱包支付
- ✓ 根据最高减扣金额自动减扣



天府通APP



2022年3月，该措施正式实施。已有6800余名教职工完成申请，成为四川大学**利用数据共享为广大教职工提供社会化服务**的一个成功案例。

不仅让教职工享受到更加方便快捷的地铁出行服务，更是学校积极响应国家**绿色出行**、低碳出行等环保理念的落地落实。



四川大学与成都天府通合作，以数据共享为基础，为智慧出行领域赋能，进一步**推进了高校和企业的数据共享**应用与创新发展的。

由于实现了业务数据闭环，信息化办在运维数据交换过程中也没有**增加额外的工作负担**。至今未发生过任何一起数据错误的情况。

为保障数据共享及应用过程的安全性，制定相关专业性要求

数据采集安全性

在采集时严格按照数据分级分类的标准对不同级别的数据进行相应安全保障。涉及师生隐私数据的通过建立独立的项目空间，授权特定的人员采集和管理；涉及少量表级的隐私数据通过按角色授权相应的采集/访问/管理权限；涉及字段级的隐私数据建立脱敏规则对相应的字段进行脱敏。

数据存储安全性

为保证数据存储阶段的数据安全性，规定数据采用三副本机制，单点故障不影响稳定性，保障数据不丢失；密码均采用国密算法存储至数据库。



数据管理安全性

在数据管理阶段通过数据权限管理、数据脱敏设计，保障校情数据的安全性。

数据共享安全性

数据中台对外提供服务时，出于数据安全的考量，基本都采用API接口方式实现。

DATA

数据**越多**价值越大，**越分享**价值越大，**越不同**价值越大，**越跨界**价值越大。

- 大力实施数据开放共享，优化治理基础数据库，不断完善数据权属界定、开放共享、交易流通等标准和措施，促使数据资产重复使用、共同使用、永久使用，加快推动跨界数据共享交换，十分必要。

数据治理是学校管理体制机制的再造，需久久为功

扩大数据的
“量”

入：数据不够全、不够完整
出：改造现有业务系统数据流程



上界

全域数据

数据交换 高基表 决策支持

下界

提高数据的
“质”

用起来（考核、个人数据门户）
准起来（一张表）
禁止私拉乱接、统一流程和标准



- 根据规则（算法）完成数据清理
- 异常数据自动探测
- 建立标签生成规则，简化数据表达

控制数据的
“权”

数据分级分类的标准与体系
确权：所有权、使用权、管理权



- 多源、异构、复杂数据的融合
- 新的数据应用，提升数据价值
- 共建共享、一次建设、重复使用

增强数据的
“智”

数据挖掘与智能决策支持
工具：AI算法、智能推荐、知识图谱



1

数据本体

数据术语集：对于共享数据对象的明确而又详细的说明

2

数据资源标识

对数据资源进行唯一性命名和主要访问手段的描述标准

3

数据共享协议

跨系统数据共享的“数据+用户”分级分类标准与协议

4

数据 → 数智

同一枚硬币的两面：大数据和人工智能，数据知识双驱



谢谢！请批评指正