

高校数据治理和服务思考

教育部教育管理信息化专家

上海市教委教育信息化2.0专家组组长

复旦大学继续教育学院院长

中国高教学会继续教育分会副理事长

宓咏 (yongmi@fudan.edu.cn)



提纲

- 背景——何为数据治理
- 要点——治理体系核心
- 实践——数据应用服务
- 探索——未来发展思考



发展历程

2000年前后，国内高校除了继续在网络、计算等方面不断优化、升级外，逐步加大了信息化应用软件的开发，高校信息化进入了一个新时期，以清华大学、北京大学、复旦大学、中国人民大学、中山大学、浙江大学、华东师范大学等为代表的一批高校开始考虑从过去独立的MIS系统建设向数据互通、认证统一的方向发展问题

复旦规划和实施原则

- 硬件**集**群
 - 统一规划系统硬件平台，集中管理，优化利用资源，节省投资
- 数据**集**中
 - 统一数据库/数据仓库，集中管理数据，存储安全、信息集中、高效利用
- 应用**集**成
 - 统一身份认证、统一应用界面风格、统一数据交换标准、统一的开发规范

基于统一的应用构建平台快速配置业务系统

教育信息化数据应用、管理服务的困境

➤ 局限性

- 面向业务条线、事务处理缺乏整合，系统基本是半独立、紧耦合
- 存在部分信息孤岛，数据质量和应用有效性不高，数据利用困难

➤ 终端用户使用遇到困难

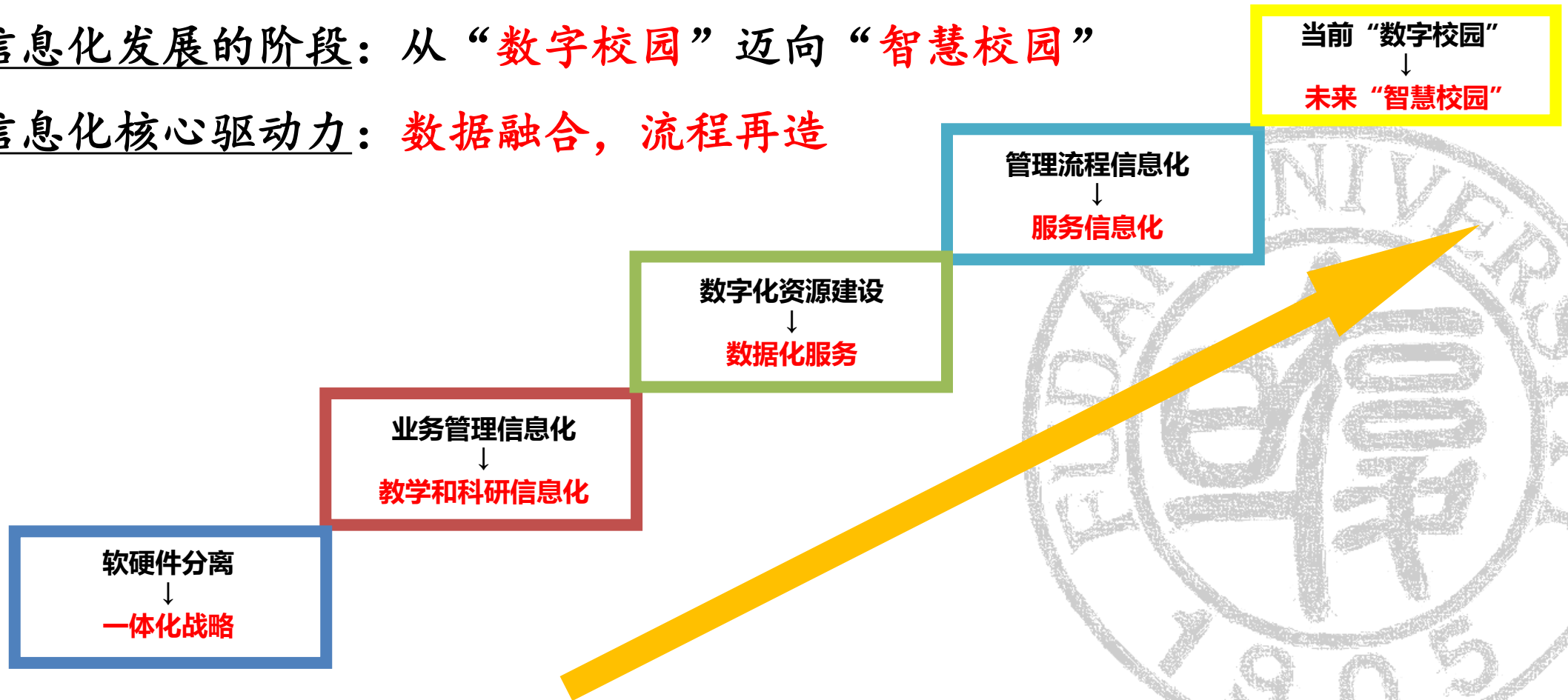
- 用户与信息库间存在信息获取、信息处理和信息提供等方面巨大矛盾，师生用户始终游离于数据产生/交换和业务贯通过程外
- 与“人”相关数据就越多、用户需登录系统越多，要全面获取数据和跨部门服务难度也越大

教育信息化发展和转变的特点与趋势

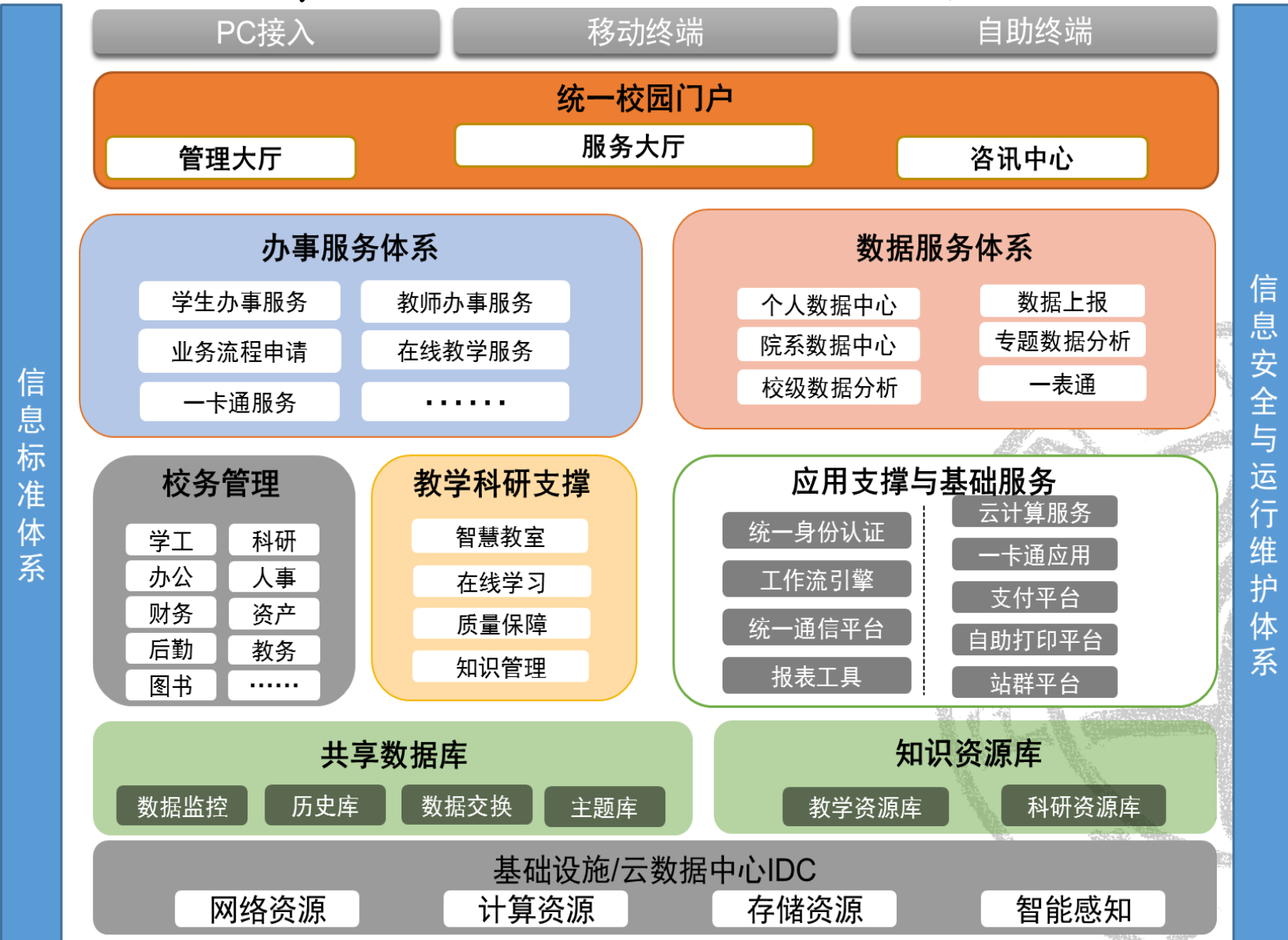
教育信息化发展的趋势：从单纯的**技术应用**向促进大学改革的深层次目标发展

教育信息化发展的阶段：从“**数字校园**”迈向“**智慧校园**”

教育信息化核心驱动力：**数据融合，流程再造**



智慧校园的共性方案

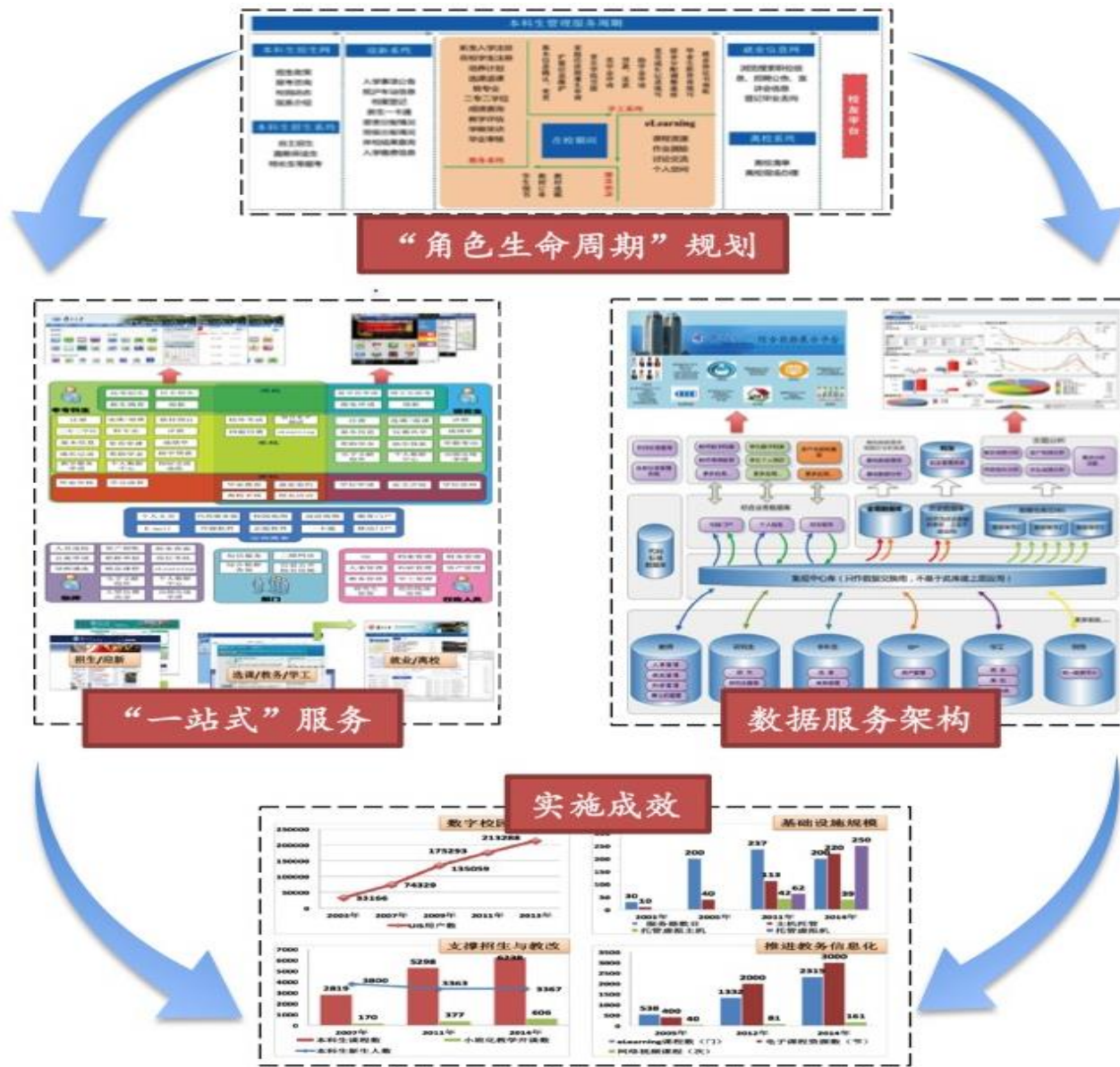


信息标准体系

信息安全与运行维护体系

重视数据、提升服务——智慧之必由之路

一站式服务
↓
流程再造
推进互联网+校务



数据化管理
↓
精准推送
实现个性化的服务



智慧校园需要完成的“三个转型”

数据化

- 1、传统数字化过程积累大量业务数据，业务部门意识不强，利用不足
- 3、目标：将非结构化数据纳入管理
- 4、形成服务成效数据化、业务模式数据化、用户行为数据化、用户属性数据化等实体

规范

服务化

- 1、信息系统以“管理”作为应用触点，梳理业务、简化优化流程
- 2、目标：以“服务”视角重新组织和整合业务过程
- 3、切断服务前端与业务后端的直接联系，释放用户的体验

精准

智慧化

- 1、完善综合业务网和物联网的融合与互通
- 2、以智能技术使得学习、资源、交流等流程贯通，推进“智能化”发展
- 3、强调对数据的综合利用和实时性，要求服务碎片化、全面性渗透到教学、科研和管理的每个角落

主动

完成**数据化过程**是创新精准服务的基石：

在信息化建设中，基础数据平台和共享库是存储数据的核心载体，保存来自各个管理系统的职能数据在数据化过程中，这些“原子”数据只有通过有效的组织，才能够形成管理者比较关心的数据化形式

充分发挥新技术作用

➤ 新技术日新月异

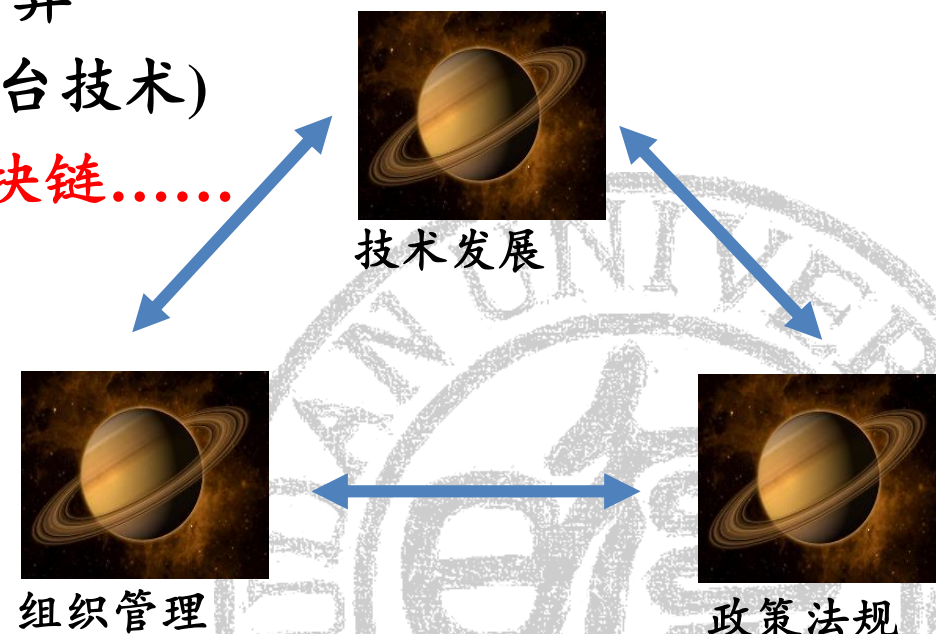
- IPv6、5G、人脸识别、无人机、边缘计算
- 微服务(一站式)、自助打印、平台化(中台技术)
- 软件架构(SaaS)、大数据、虚拟化、**区块链.....**

➤ 适应、协同

- 技术发展：一日千里
- 组织管理：一日百里
- 政策法规：一日十里

➤ 技术的作用

- 传统模式的电子化 -- 工具
- 新技术环境下的模式转型成为 -- 催化剂、助推剂



切实理解互联网+

“+互联网”与“互联网+”之间的区别

- “+互联网”是“各传统业务+互联网”，只是让信息技术服务于传统业务，提供便捷的服务；而“互联网+”是“互联网+各传统业务”，则是利用互联网平台让各业务深度融合起来，扩大服务范围，提高服务质量，帮助业务发展，最终实践信息化转型之路

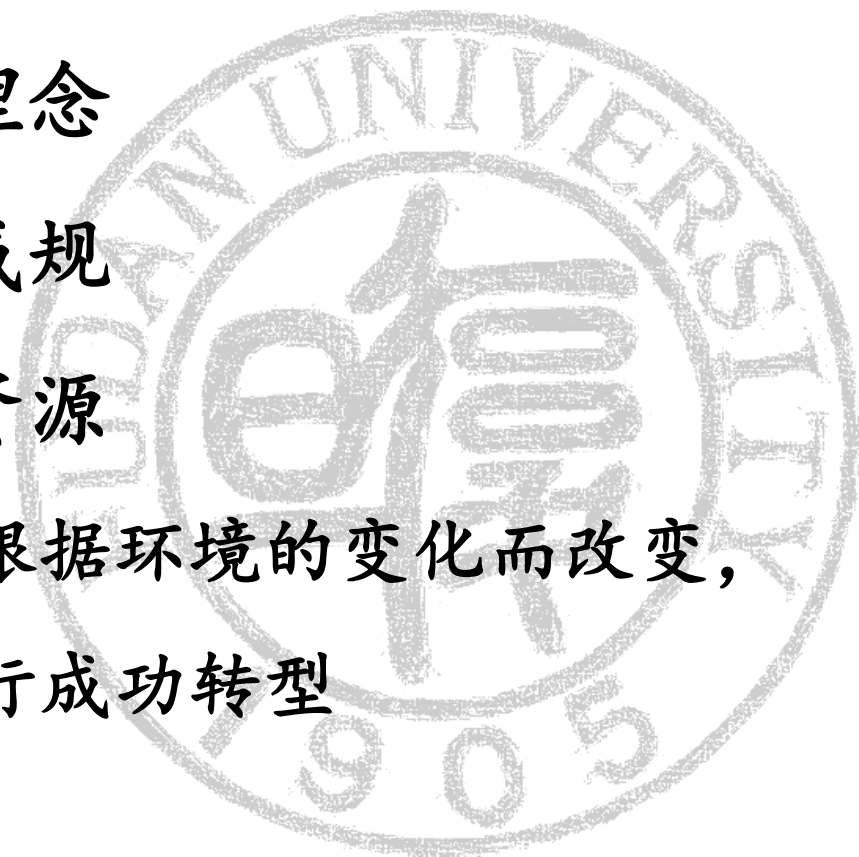
“口碑” → 互联网时代信息化评价的核心指标
最后一公里问题：建设初衷与用户体验**严重反差**
如何从“能办”到“好办”、最终到“愿办”？

需要从信息化支持升级到信息化赋能！

互联网思维与传统思维的五大区别

- 互动思维 ~ 注重互动，而不是单向传播
- 免费思维 ~ 推崇免费，而不是强调收费
- 平等思维 ~ 强调平等，而不是层级理念
- 开放思维 ~ 彼此共生，而不是墨守成规
- 整合思维 ~ 懂得整合，而不是分散资源

传统思维模式在当前互联网时代中，必须根据环境的变化而改变，
跟上时代步伐、完成思维转变，才能最终实行成功转型



教育信息化转型中的要素

➤ 服务治理

- 制定规范、搭建平台、接纳服务

➤ 能力开放

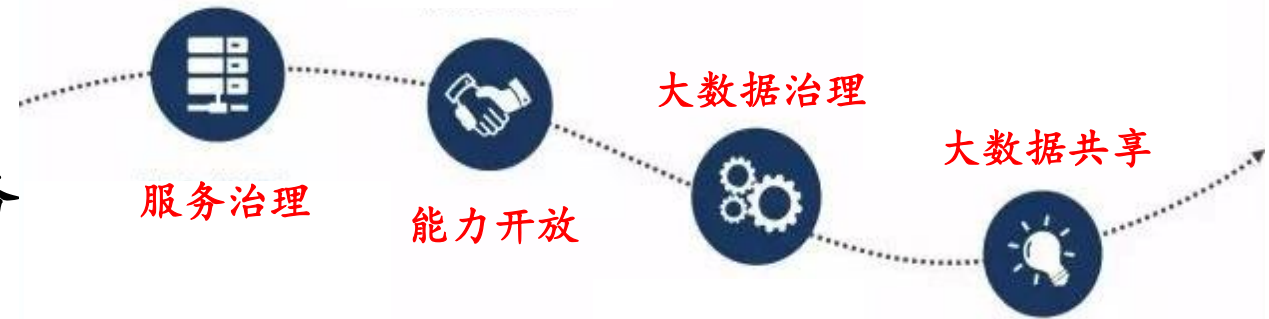
- 能力开放平台---为终端用户提供更广泛的接触渠道，提高沟通效率

➤ (大)数据治理

- 数据资产管理平台---提高针对全生命周期的治理能力，为给予大数据的应用创新提供基础

➤ (大)数据共享

- 数据共享平台---在数据可靠的前提下为其它合作伙伴和互联网个人开发者提供数据共享开放服务



教育改革核心要素

➤ 在技术上实现管理体系的“去行政化”，以顺应从“管理”走向“治理”的趋势

➤ **服务治理** —— 行政服务规划入手，梳理与整合业务流程，并再造、以提供更明晰化的服务

➤ **数据治理** —— 建立数据资产管理平台，提高治理能力，为基于大数据的应用创新提供基础

管理与治理

➤ **管理**是由上而下的，**治理**是全生态的

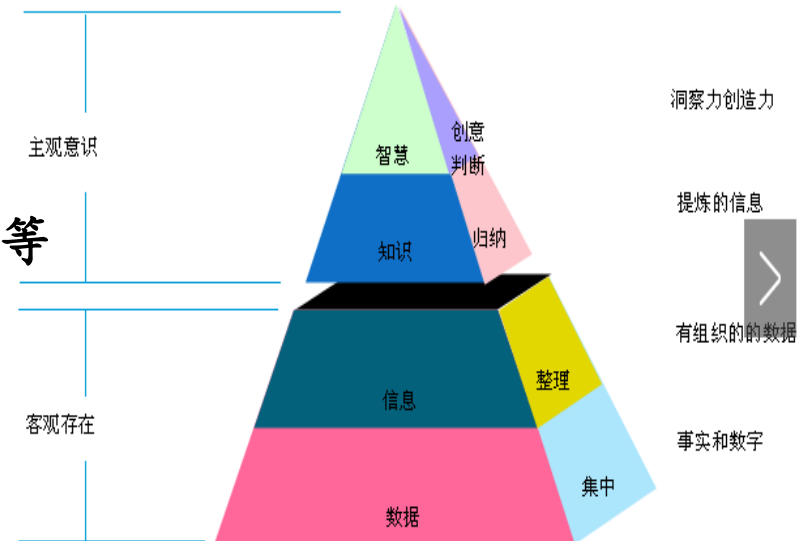
➤ **管理**是被动的，**治理**是主动的

➤ **管理**针对行政业务，**治理**面向全面、全方位的工作



数据是信息化的核心资源

- 多年信息化建设和应用已积累了大量教学、科研、运行管理等方面业务数据，传统的以流程为主体的业务服务模式所沉淀下来的数据成为了新的宝贵资源
- 数据→信息→知识的智能演化路径强调对数据的深层次利用，如何让数据服务于学校治理是我们目标
- 信息化发展进入新阶段--智慧校园、云服务、泛在学习、知识管理.....
- 信息化数据的形态也在发生变化
 - **传统文档、资料的再加工** — 资源“知识化”
 - **非结构化数据重要性凸显** — 教学资源、视频资料、科研数据等
 - **WIFI、物联网等新技术普及使得大量微观数据触手可得**
 - 地理信息、用户行为等
- **数据反哺** —> 提升服务水平、提高管理效率



数据的价值

业务用户

希望提高业务部门的效率



数据管理者

需可信数据支持快速工作



行政部门

实时、灵活的调整相关策略



领导层

全面精准的了解单位状态



管理者

希望提高IT项目的投资回报率



运维部门

希望能够改善并减少用户意见



信息化业务数据分类

业务数据分类有多重维度，根据传统信息化的业务领域可划分

➤ 业务数据

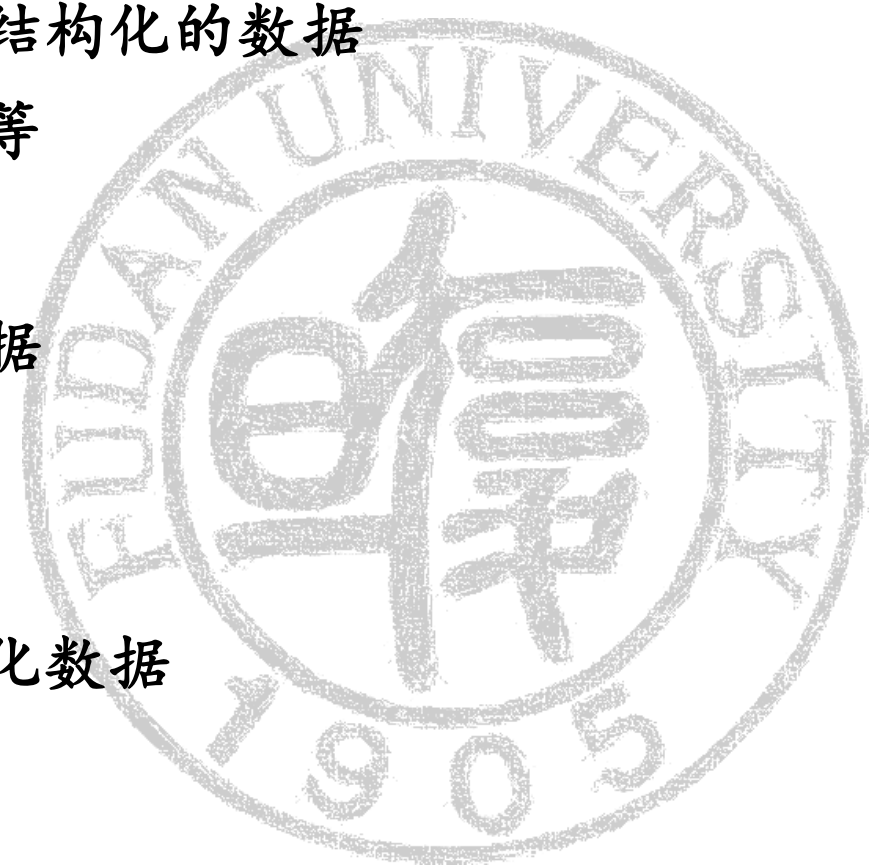
- 业务系统、一卡通和服务应用等产生可获得的、结构化的数据
- 如人事基本信息、学生选课和成绩、一卡通消费等

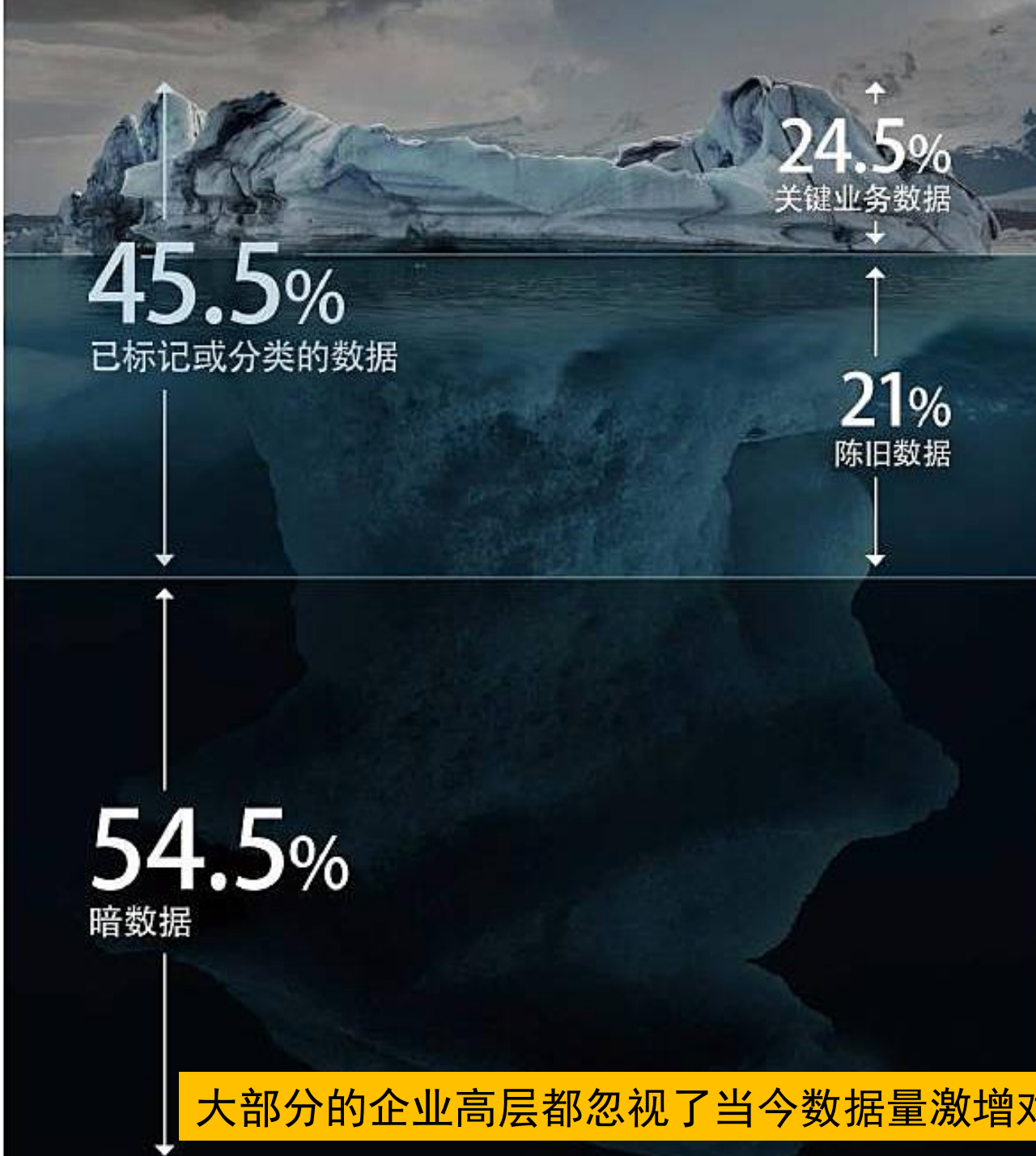
➤ 网络数据

- 网络基础设施和基础网络应用生成的用户行为数据
- 比如校园网用户使用信息、WIFI和物联网数据等

➤ 其他数据

- 未纳入统一管理的数据、难以直接利用的非结构化数据
- 比如网络教学资源、视频资料、科研数据等





- 关键业务数据：该类数据是企业持续成功运营的关键；**企业不但需要妥善保护该类数据，更需要主动对其进行管理**
- 冗余、过期和不重要的数据：该类数据为冗余、重复、过期、没有商业价值的的数据，以及仅有少量商业价值的非关键数据；**企业应该主动地定期删除陈旧数据，将该类数据控制在最低水平**
- 暗数据：该类数据的价值尚不明确，但无论是否具有价值，暗数据都会消耗一定的企业资源。企业需要尽早了解暗数据的构成，并对数据进行相应的划分——**关键业务数据 或 陈旧数据；最重要的是，暗数据可能包含高风险的非法数据，导致企业IT系统的核心存在严重的潜在风险**

大部分的企业高层都忽视了当今数据量激增对企业所带来的影响

来源：《数据冰山报告》

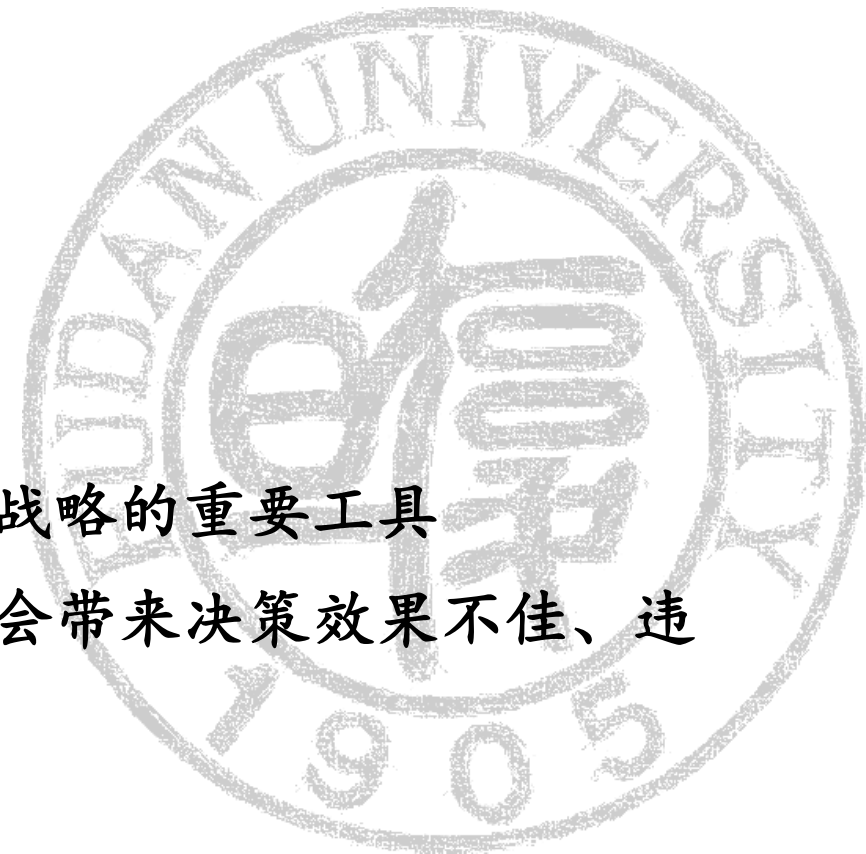
数据治理

数据治理是指从使用零散数据变为使用统一主数据、从具有很少或没有组织和流程治理到单位范围内的综合数据治理、从尝试处理主数据混乱状况到主数据井井有条的一个过程

- 提升数据可用性、数据质量、数据隐私和数据安全为主要目标的基础工作
- 信息系统建设发展到一定阶段，数据资源将成为战略资产，而有效的数据治理才是数据资产形成的必要条件
- 对于确保数据的准确、适度分享和保护至关重要，有效的数据治理计划会通过改进决策、缩减成本、降低风险和提高安全合规等方式，将价值回馈于业务

数据资产和管理理念

- 数据是单位/部门最核心的资产
 - 带来利益和效益的资源
 - 兼有无形资产和数据资源的双重特征
 - 需要作为无形、有价值、核心的资产进行管理
- 数据资产的价值体现
 - 重要资源---是赖以运转的重要资源
 - 创造价值---创造价值的最大来源
 - 决策方法---数据驱动下的决策方法是实施发展战略的重要工具
 - 动态发展---对动态变化估计不足、管理不善，会带来决策效果不佳、违规和数据失窃等



数据服务化

- 随着数字化的转型，积累的大量数据，正在形成丰富的数据资产金矿，但在**有价值的数据**和**数据产生价值**之间仍然存在“最后一公里”的跨越；但这就像要打通断头路、非常难，通常要涉及多个部门的密切配合才能搞定；但一旦打通，带来的效果也是非常显著的
 - 当前应用的三大特征：业务推出快、应用粒度散、API化和服务化
 - 从关注技术到关注数据，回归数据应用价值本身
 - 数据处理由离散工具向数据服务转型，建设面向服务共享的数据平台
 - 数据服务共享平台定位：数据资源“纵向贯通”、“横向互联”的共享通道，成为数据运营中心、成为心脏
 - 通过数据服务平台，降低业务对技术依赖，充分发挥业务创新潜能，打造数据资产的运营能力

形成数据服务的生态环境

- 面向数据服务**提供者**和数据服务**消费者**，梳理清晰整个服务化流程
 - 数据服务提供者：完成数据接入，服务管理，消费方管理、运行监控能力
 - 数据服务消费者：通过服务申请来完成对服务的消费使用，并提供权限范围内数据浏览与查看
- 顶层设计统一运营数据，实现服务开放
 - 统一的数据共享通道：建立统一的纵向数据共享通道，提供跨系统、跨部门、跨区域的数据集成、交换、分发、共享机制
 - 统一的数据交换标准：定义统一的交换标准和规范，并在实施过程中，积累更多符合自身发展的、可复制的最佳实践
 - 统一调度管理：提供灵活的、多角度的模型作业调度机制，减轻运维管理工作量，实现运维自动化
 - 统一数据运行监控和安全保障体系：提供数据服务发布、访问授权和运行监控的统一管理；从技术和管理两方面提供事前预防、事中控制和事后追溯能力

提纲

- 背景——何为数据治理
- 要点——治理体系核心
- 实践——数据应用服务
- 探索——未来发展思考



数据治理规划的要素

数据治理是一个系统体系，是一个从上至下指导、从下而上推进的工作；要达成共识、有一个强有力的组织、合理的章程、明确的流程、健壮的系统，来有效保障数据治理工作

- 发展战略目标---提供数据服务、体现数据价值
- 数据治理组织---管理团队的架构、层次、职责
- 数据治理制度---政策、制度、细则、规范体系
- 数据治理流程---流程管理的目标、任务、分级
- 数据治理技术---资产、质量、挖掘和互联互通



强化数据治理

➤ Data Governance(御数坊CEO)

- governance ≠ management, governance ≠ framework
- governance = management of management

➤ 数据生态治理是长期工作，而非一蹴而就工程

➤ 是帮助我们理清每条数据分工和责任人的体系

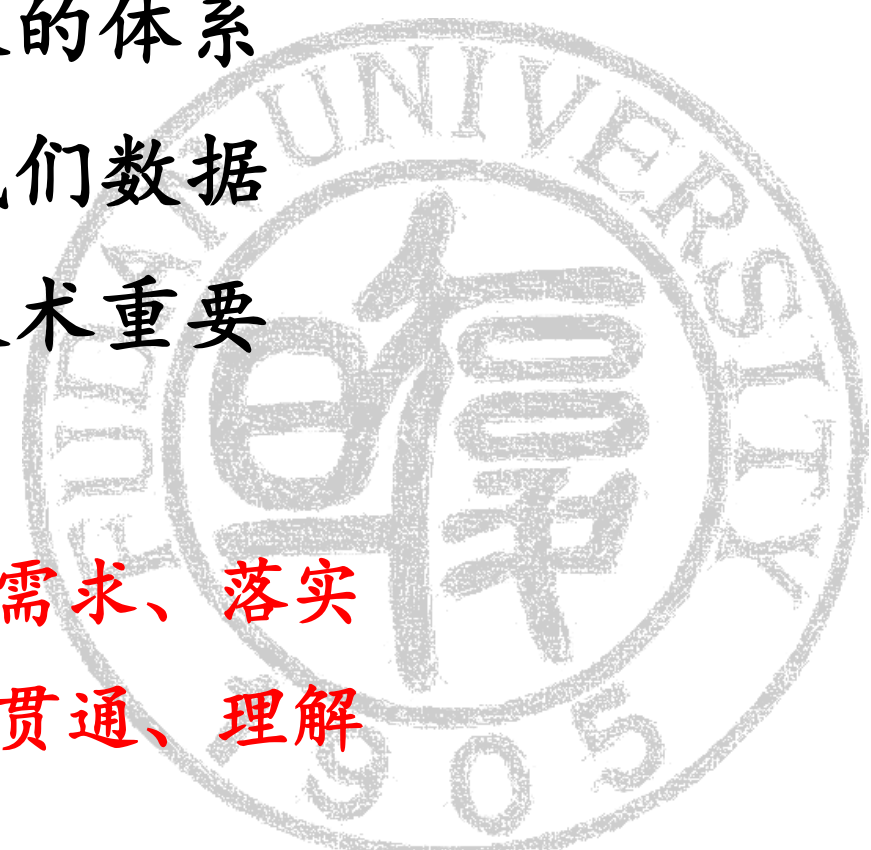
➤ 应该像管理财务那样，去妥善管理好我们数据

➤ 数据资产是权力、也是责任，需求比技术重要

➤ 数据治理的两个维度

➤ 垂直：决策层、业务层和实施层落地 → 需求、落实

➤ 水平：业务、数据和技术三部门配合 → 贯通、理解



数据治理的落地

- 规范管理数据资产理念已被广泛接受，但还需强化组织架构、原则、过程和规则，确保数据管理的各项职能得到正确履行
- 数据治理是一个关注于信息系统执行层面的体系，其目的是整合IT与业务部门的知识和意见，通过虚拟组织对信息化建设进行全方位监管，因此需要高层授权、业务部门与IT部门建设性的合作
 - 数据治理范围涵盖了从前端事务处理系统、后端业务数据库到终端的数据分析，从源头到终端再回到源头形成一个**闭环负反馈系统**
 - 数据治理要对数据的**获取、处理、使用、服务、反馈**进行监管，并通过**发现、监督、控制、沟通、整合**等五方面执行力来落实和保障

何为教育信息化数据治理体系？

- 目前教育信息化数据管理的各个环节存在不少问题，信息化数据治理体系是结合信息化实践、形成的总体解决方案
- 教育信息化数据治理体系的核心内容，是利用信息化技术，将学校业务数据通过**采集、处理、利用加以高效组织和科学管理**，使其成果得以有效地应用于学校的教学、科研和校务管理与决策过程，使其有效提升学校的核心竞争力

高校数据体系的特点

- 与企业相比，学校的数据生产和利用过程存在以下特点——数据来源分散、数据数量众多、数据生产机构业务复杂多变、技术部门强制力不足、数据分析需求层次错落且目标不够清晰……
- 一言概之——学校因其自身业务特点，**注定信息化部门不能等同于(难以拥有)企业IT部门的管理职能**；学校信息化部门更多定位于服务职能，难以通过强制手段获取并为单一、清晰的数据利用目标服务
- 因此，建立数据生态，探索一种有效、便捷的数据生产和利用模式(不同于企业的强制管理模式)，是教育信息化数据工作的核心内容

教育数据治理和服务体系



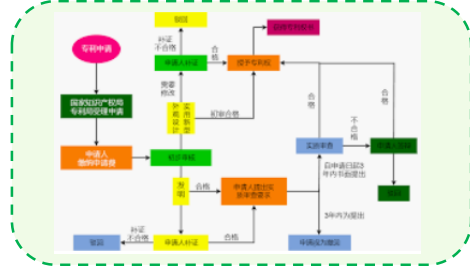
数据治理体系

组织：数据治理团队



- 数据委员会
- 数据所有者
- 数据管理员
- 业务专家
- 质量分析师
- 开发人员
- 数据架构师
- ...

标准、流程、策略



定义

- 数据定义
- 技术标准
- 企业数据模型
- 主数据/引证数据
- 业务规则
- 数据安全
- 数据访问和管理
- 数据监控和审计

实施

集成技术平台



循环式开发和交付

- 数据实时访问
- 数据转换
- 模糊匹配和整合
- 数据交付
- 数据标准化

提纲

- 背景——何为数据治理
- 要点——治理体系核心
- 实践——数据应用服务
- 探索——未来发展思考



数据服务体系规划



- 内容：建立面向高校的创新型层次化数据服务体系，覆盖信息化所有应用主体
- 目标：以个人、二级单位和学校为主体，建立数据采集、流通和服务机制

数据生态核心要数—数据的生命周期

多层次角色主体(hierarchical roles)成为数据消费的受益者，提高数据生产积极性

重点关注和研究数据生命周期四要素

从哪来?

数据从“私有”为“公有”，
纳入大量游离在外的数据

怎么来?

变技术主导为需求主导，
将数据真正流转到业务
专家手中、产生价值

降低数据收集成本，数据
获取工作独立于业务系统

到哪去?

重新加工、数据服务独
立于业务系统，数据服
务业务和技术同等重要

如何用?

将各级用户角色纳入数据生态体系，使之成为数据消费的受益者，提高数据生产积极性，推动整个数据生态的有效运行



规划层次化特点的数据服务体系

三层结构

师生个人数据中心

教师/学生个人**查询、填报**

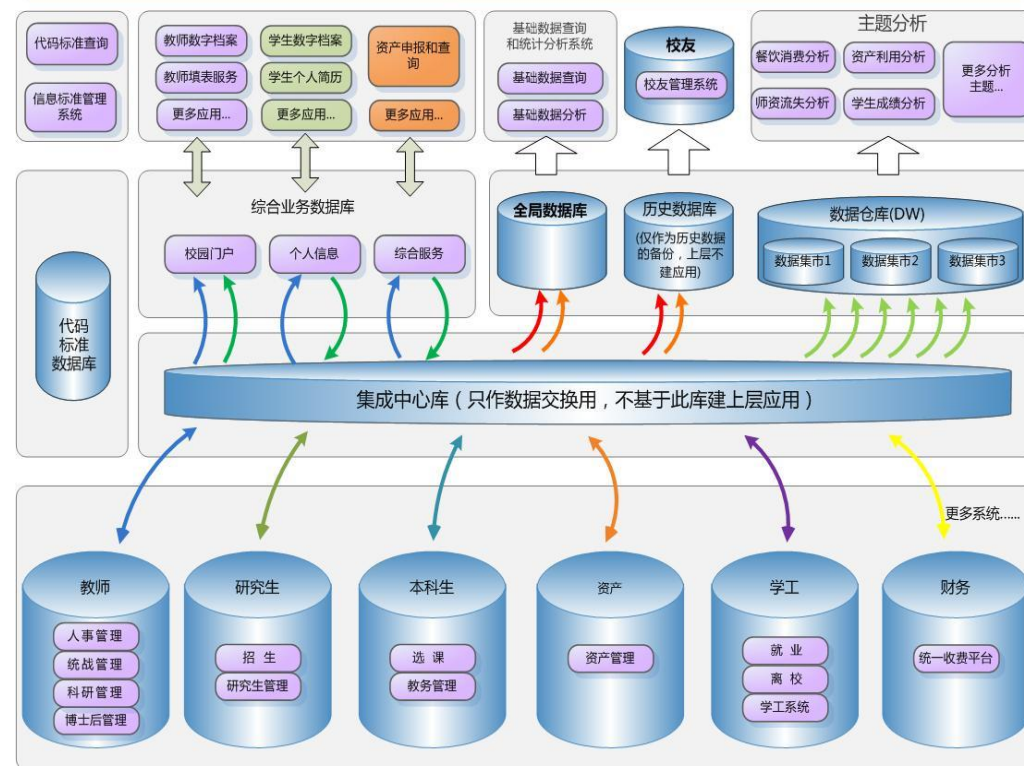
院系二级部门数据中心及分析平台

院系/部门**快速、灵活配置**

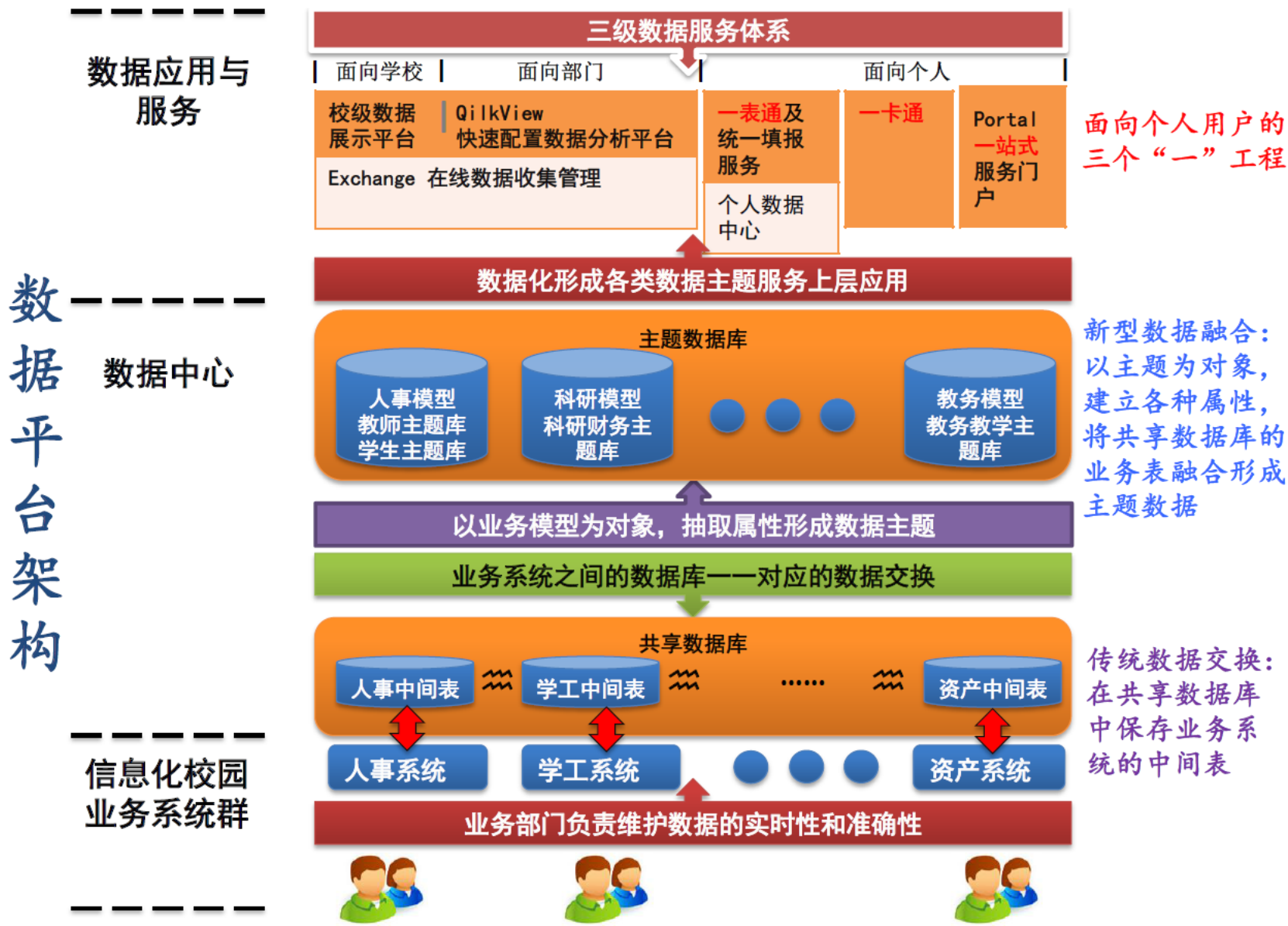
校级数据展示和决策支持平台

校级**分析展示、决策支持**

完善信息化校园数据架构



数据服务整体架构



调整数据组织视角

学校视角

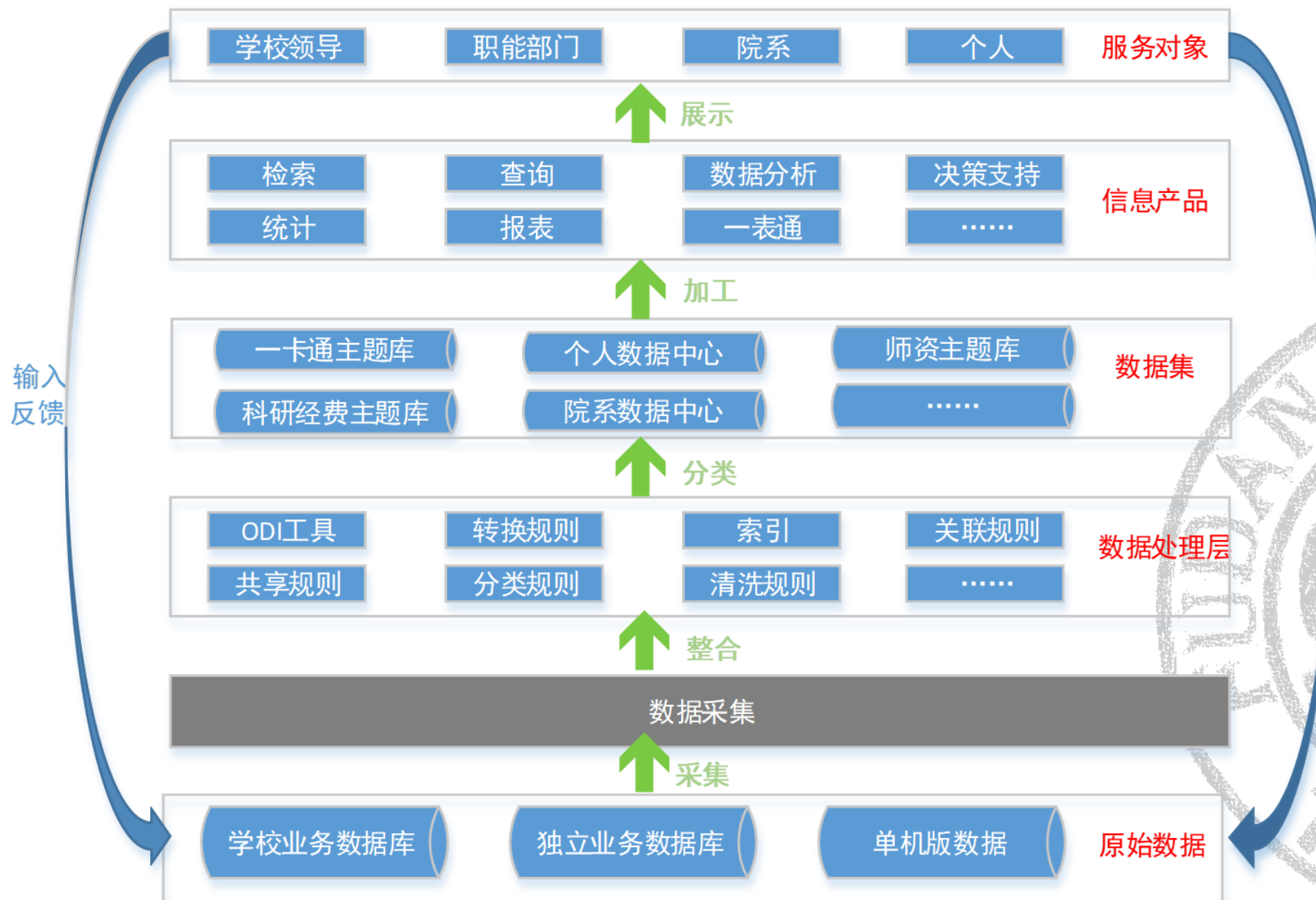
院系视角

个人视角

人员信息	论文信息	财务信息	项目信息	资产信息	教学信息
年龄	标题	总额	类别	设备	学生信息
学位	排名	余额	归属	型号	课程列表
论文	影响因子	拨款单位	余额	购置年月	奖学金
.....
人事系统	图书馆统计	财务系统	科研系统	资产系统	教学系统

数据打通和重组

建立数据良性循环机制/生态环境



逐步形成数据产生、维护、应用、受益的良性循环机制和生态环境

使各环节用户成为数据的生产者、维护者、使用者和受益者

数据平台建设架构

- 一. 建立面向多源异构数据、尤其是非结构化数据的自动采集机制
 - 数据来源可能是合作方或外部热点站点，需针对不同类型数据定制化开发不同的采集程序，如包装器、抽取程序等
- 二. 文本、图像、视频等非结构化数据，必须与人员、组织、产品等主数据进行关联
 - 通过索引、分析等技术，挖掘非结构化数据的潜在价值
- 三. 架构应逐渐向“以机器为学习主体”目标架构阶段性演进
 - 与传统数据仓库建设方法类似：先根据应用需求，选择知识图谱、智能机器人等应用方向，基于特定领域设计应用场景，逐步建立起针对内外部数据自动采集、整合、分析以及自动学习优化算法的能力；然后通过场景驱动逐渐推动大数据应用架构向大数据智能化架构演进

面向对象的业务数据整合、重组

科学规划，重构数据基础，以“面向对象”方式组织教学、科研、财务、人事、生活等各类数据资源

数据管理后端和前端服务分离，实现“**松耦合**”

- 数据的交换、保管等过程，由**信息化部门**统一操作
- 数据产生(收集)和校验由**用户与二级单位**共同承担
- 数据能被“一站式”服务等**直接、便捷、快速**调用
- 分析和应用由**二级单位与学校**灵活定义和整理利用

需关注的因素

覆盖生命周期 全系列数据

- **学生服务管理**从招生、迎新，到在校期间的日常管理，再到毕业、学位、就业、离校等环节
- **教师服务管理**的校园电子公务、人事管理、组织管理、科研管理、教学管理、资产管理等系统也都经过了多年的实践、运行

研究技术支撑 的基础要素

- **平台基础支撑** -- 数据标准、存储、计算、处理和安全机制
- **数据融合机制** -- 数据采集、清洗、校验、抽取和形成主题机制
- **数据服务应用** -- 一站式、数据展示分析等数据服务

解决学校关心 的核心问题

- 支撑教学改革
- 改变管理理念
- 提高管理效率
- 提升服务水平
- 增强核心竞争力
-



基础数据平台的设计和实施

➤ 数据产生

- 统一制订代码规范标准
- “谁产生、谁维护”
- 视图、变动跟踪表、中间表等方式保证业务系统数据变更独立、安全

➤ 数据接口

- 建立视图实现业务逻辑到数据交换逻辑的转换
- 触发器、变动记录表、时间戳等方式，个别采用跟踪日志
- “主动推”——数据及时采样
- “被动取”——数据定时采样

➤ 数据总线

- 星型分布--独立，松散耦合，易维护和升级

➤ 交换中心

- 过滤测试数据、识别非法数据
- 将系统私有代码转换为校级公用代码

➤ 主题数据

- 中心交换层已把分散在各个业务系统的数据汇总，但数据仍然按业务组织
- 将各业务系统的数据按主题聚合，简化分析
- 教职工、本科生、专科生、科研等主题

➤ 历史数据

- 高校业务具有时间周期性，业务部门通常只关心本学年、本学期数据，因此在业务数据库中积累历史数据缺乏推动力
- 在基础数据平台中积累历史数据，建立时间关联

基础数据准备

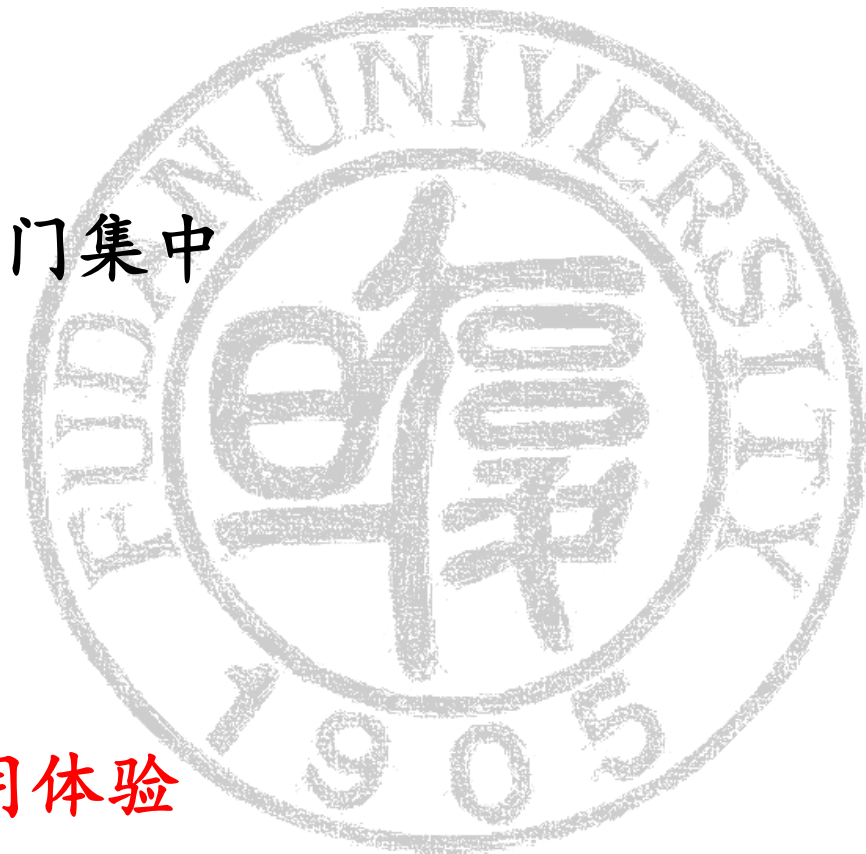
- 系统基础数据分类
 - 静态数据和动态数据
 - 显性数据和隐性数据
 - 主数据和辅助数据
- 整理基础数据方式
 - 成立数据小组，确定工作进度
 - 数据整理自上而下、先易后难
 - 把好数据录入关
- 基础数据整理工作的重点之一
 - 抓住编码这一主线



基础数据收集基本步骤

- 确定工作范围
- 建立必要的编码原则
- 建立公用信息
- 确定基础(基本)数据结构
- 收集第一手资料，将离散数据从不同部门集中
- 数据检查(完整性、正确性、唯一性)
- 数据录入应用系统
- 系统检核

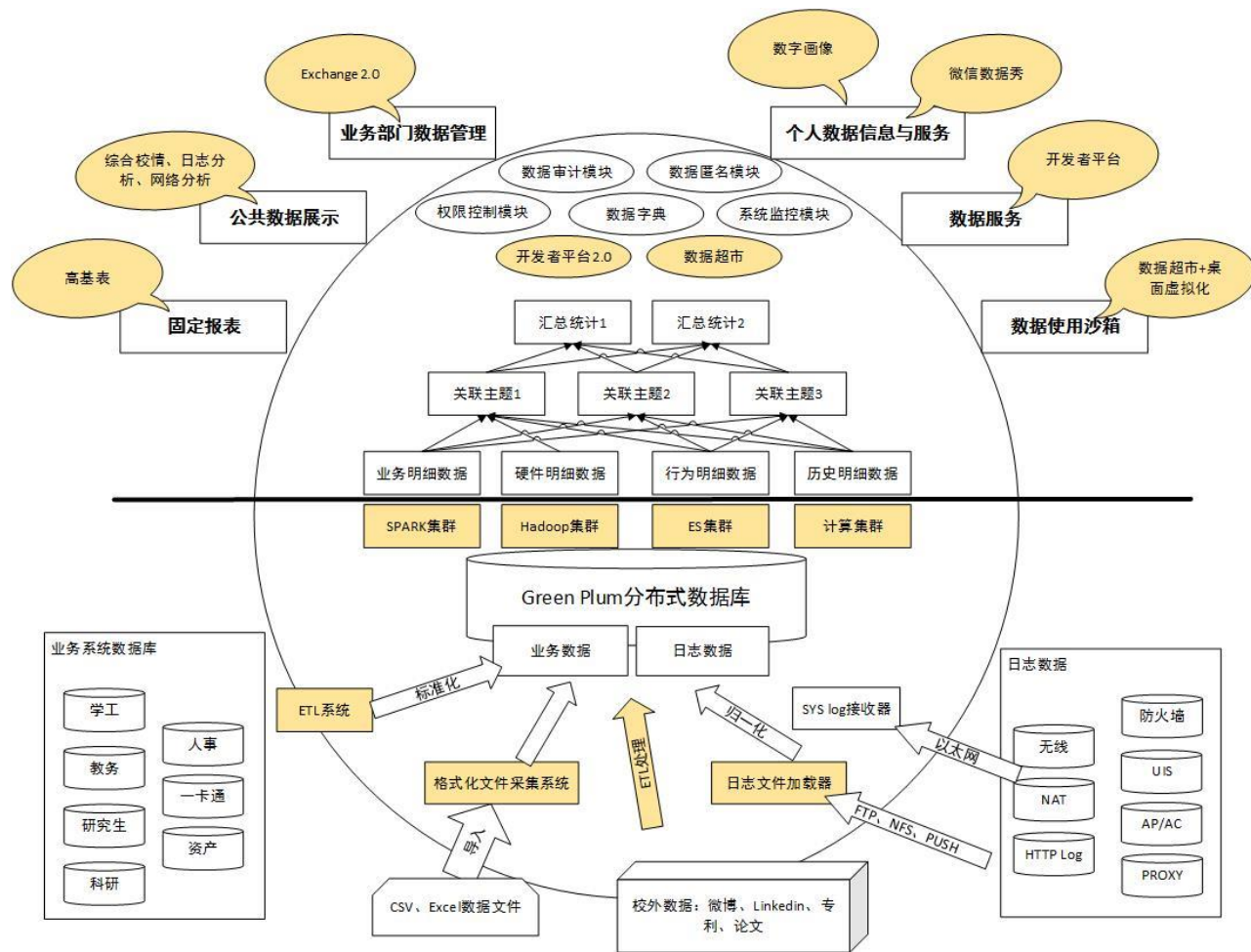
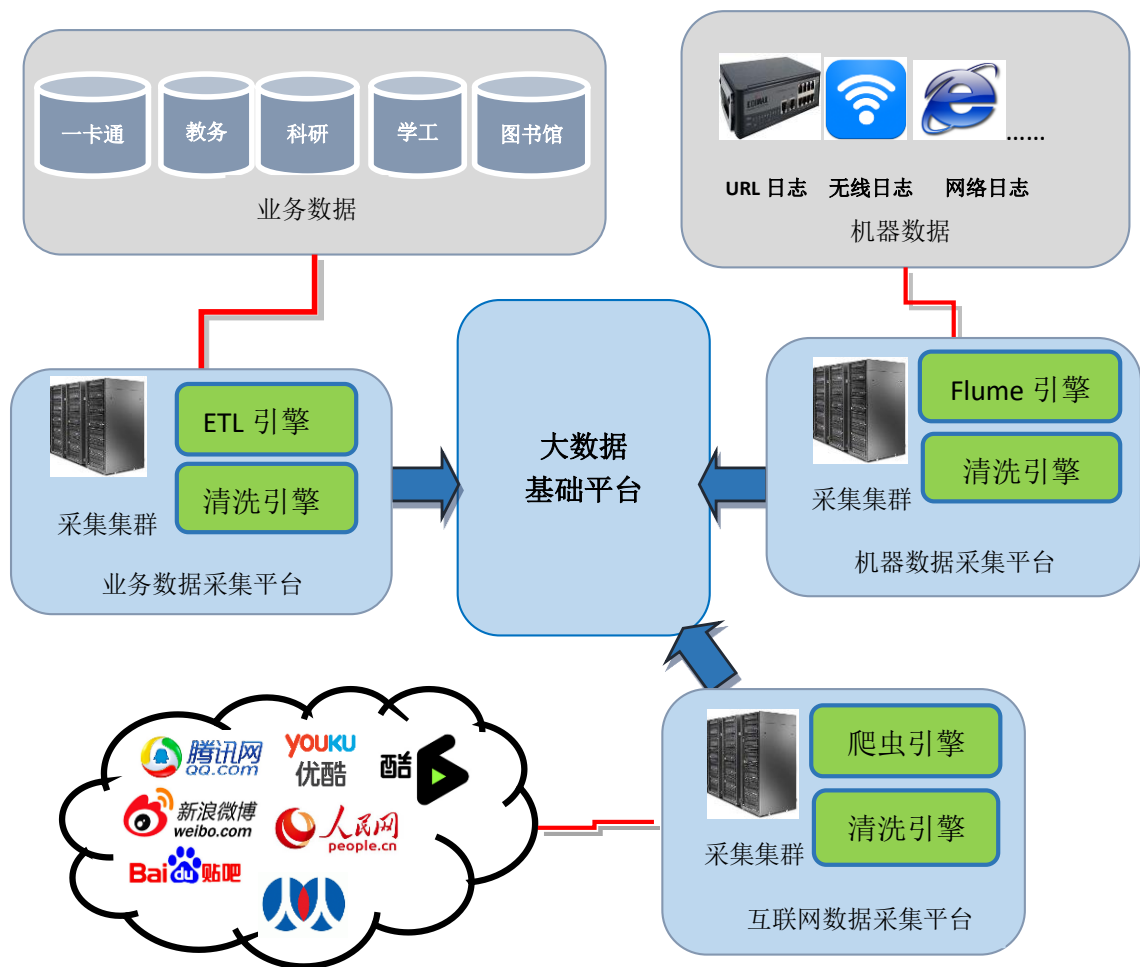
不夯实基础数据、就很难有好的信息化应用体验



快数据、大数据

- 随着智能设备不断接入网络(万物互联), 能采集的数据呈现出几何级数爆炸式增长, 如何面对海量数据面临新的挑战
 - 沉重的数据处理压力、永无满足的存储空间及存储管理能力
 - 物联网设备、人工智能技术的发展, 急需数据进行赋能
- 快数据、大数据处理成为技术瓶颈, 数据宝藏成为数据负担
 - 快数据---及时处理、迅速反馈(快速准入、持久耐用、低功耗)
 - 大数据---海量能力、非结构化(动态跨云、接口兼容、可靠性)

存储大数据-高校数据采存和处理体系



使用大数据-数据使用氛围

- 业务部门有数据治理意识
- 有效掌握数据分析的技能
- 领导有数据管理应用意识
- 教育大数据研究学术氛围

必要的软实力

- 管理机制的保障
- 专职部门行政保障
- 人才梯队的培养机制

数据质量问题的误区

——是“技术”问题还是“业务”问题

➤ 数据质量问题的危害

- 错误或不完整数据导致BI和MIS系统不能正常发挥优势
- 数据分析员每天会有超过30%的时间精力浪费在辨别是否“坏数据”上
- 低劣的数据质量严重降低了工作效率和产生的效益

➤ 数据质量问题的来源

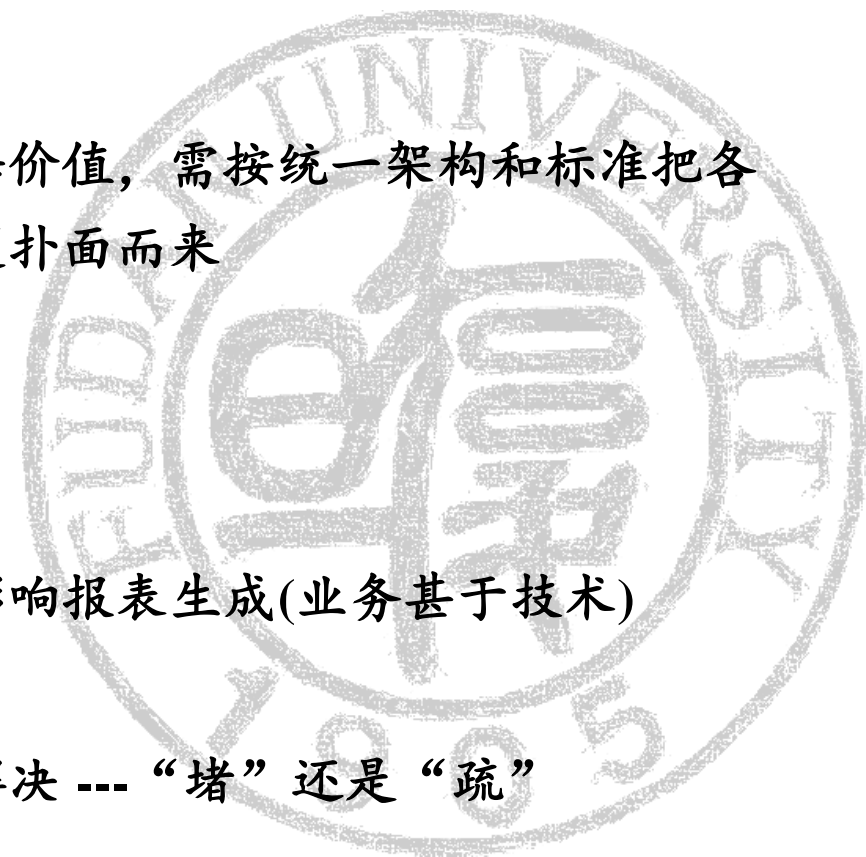
- 初期各类业务系统快速增长，随着业务需求增长和挖掘数据价值，需按统一架构和标准把各类数据集成融合，数据不一致、不完整、不准确等各种问题扑面而来

➤ 数据质量问题的责任主体

- 数据来源渠道多、责任不明确(业务)
- 业务需求不清晰、数据填报缺失(业务)
- ETL处理过程中，业务部门变更代码导致数据加工出错，影响报表生成(业务甚于技术)

➤ 解决数据质量问题

- 弄清楚数据质量问题来源、责任主体，应该仔细考虑如何解决 --- “堵”还是“疏”



解决数据质量的模式

- 采用规划顶层设计，制定统一数据架构、数据标准，设计数据质量的管理机制，建立相应的组织架构和管理制度，采用分类处理的方式持续提升数据质量，这是数据质量管理“疏”的方式
- 单纯依赖技术手段，通过增加ETL数据清洗处理逻辑的复杂度，使用数据质量工具来发现ETL数据处理中的问题属于“堵”的方式，只能解决表面的问题，不是根本的解决方法
- “堵”的方式其实也在好多单位中都在使用，其根本目的在于提高ETL处理的准确度，做法无可厚非，毕竟需求的出发点不一样，面临的现状也不一样

有效解决数据质量问题



- 定义清楚业务问题，以业务问题为出发点做顶层设计
- 建立数据质量闭环管理机制(问题的定义、发现、整改、跟踪，及效果评估)
- 对不同类数据(历史数据、当前数据、未来数据)的数据问题作分类处理、要采取不同处理方式

分类处理、采用针对性处理方式

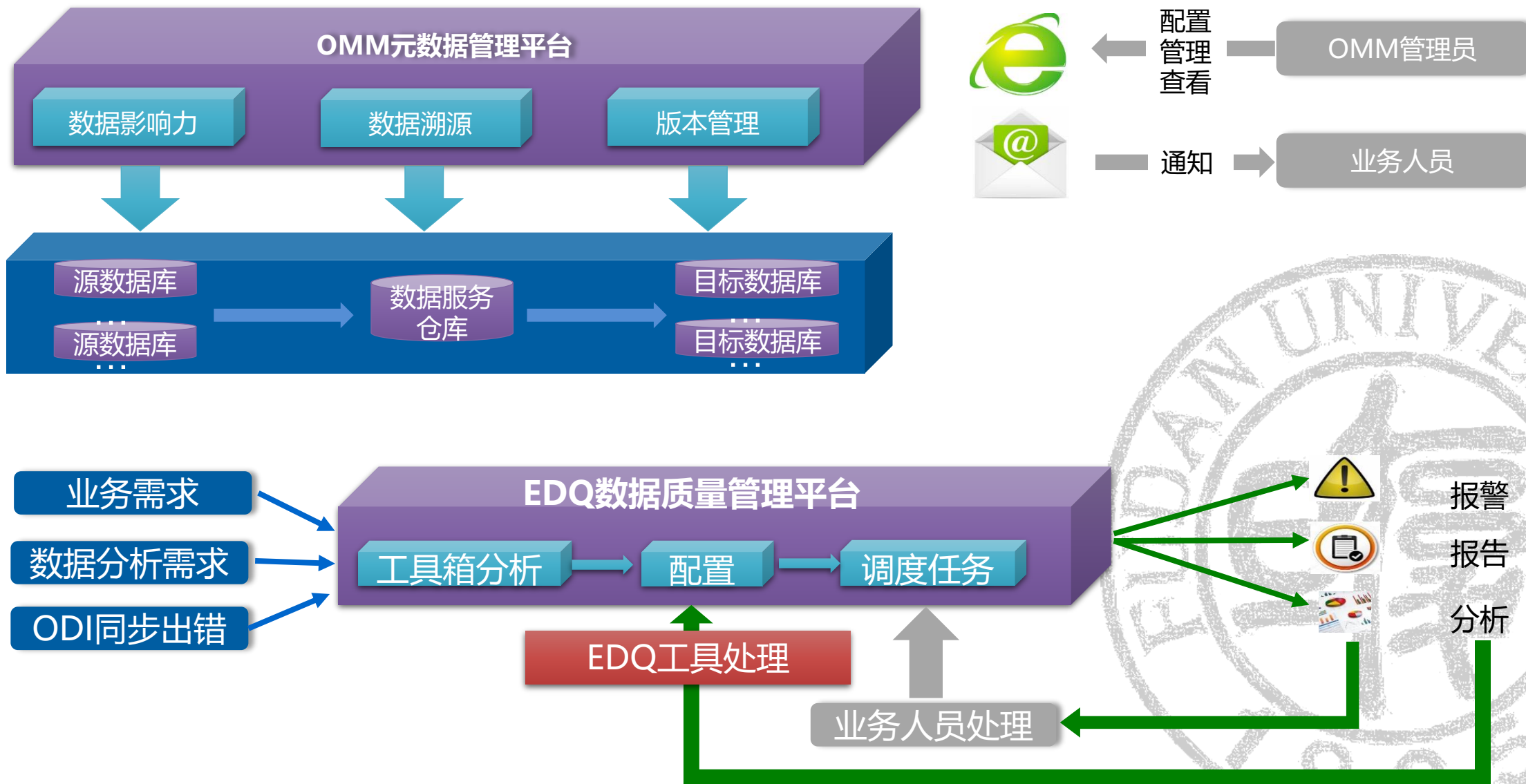


- ▶ 历史数据：发挥IT技术人员优势清洗数据，综合使用各类数据源，提升历史数据质量
- ▶ 当前数据：需要从问题的定义、发现、整改、跟踪，及效果评估5个方面来解决
- ▶ 未来数据：制定数据规划，从信息化角度出发、统一规划学校数据架构、制定数据标准和模型，借系统升级/重建之际，从根本上提高数据质量

妥善应对数据质量问题

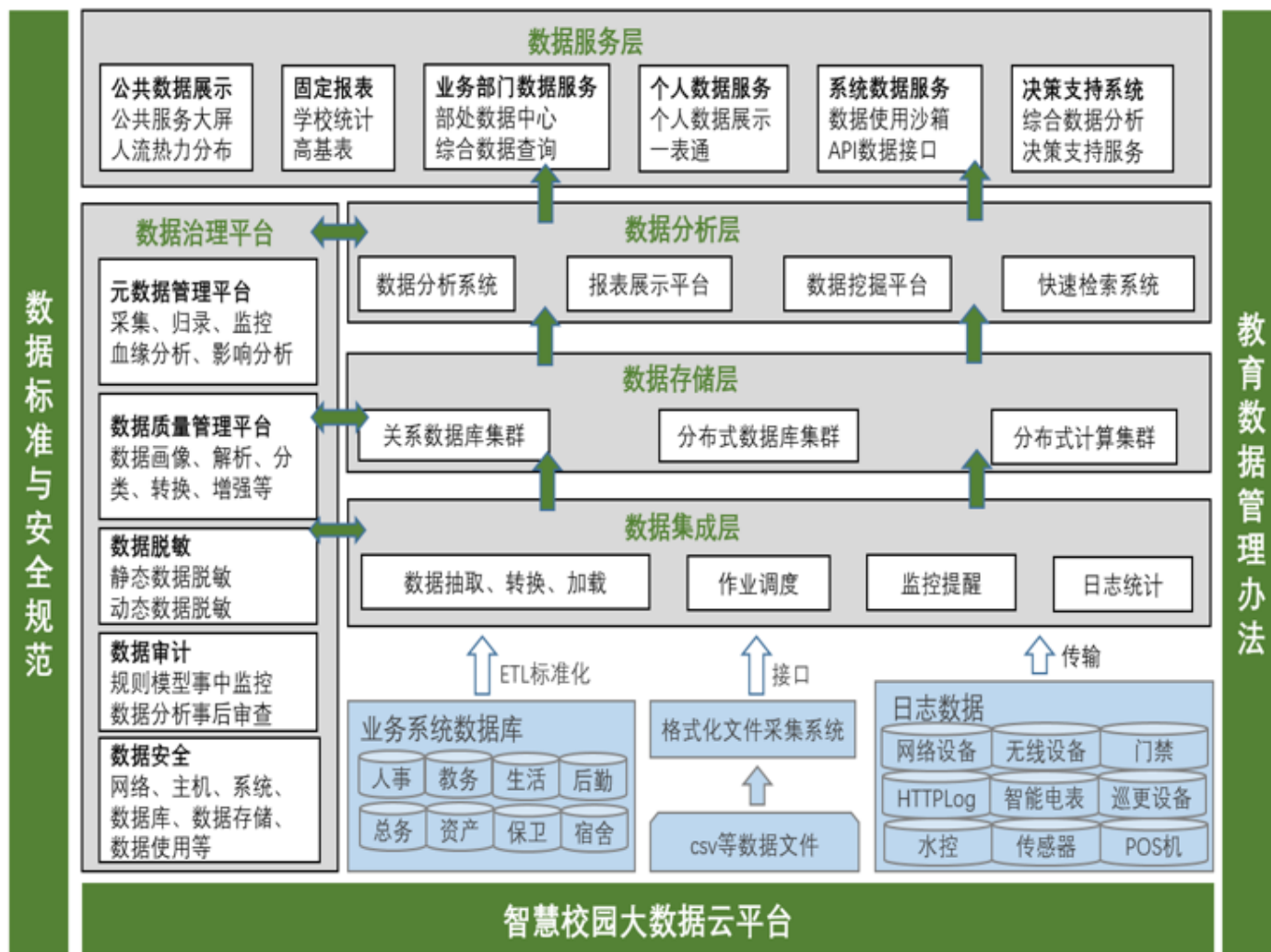
- “我们现在还没有到利用数据这一步，因为(应用系统中的)数据质量太差”——>“还没到培养孩子的时候啊，孩子太小了”
- 不能因为数据质量差、就不去利用数据，恰恰是因为没有去做后面的事情、所以数据质量才差；而且也不能因为数据质量差就抛开业务场景、试图直接全面解决数据质量的问题，这样得不到业务部门的支持，也无法从数据工作中产生业务价值。所以更要利用做应用、做业务的需求，同步解决数据质量问题
- 数据质量问题，根本上是在构建应用之初缺乏整体数据规划和数据思维导致的问题，没有从源头上优化数据的存储、流转，没有与其它系统数据去对齐口径、统一语言

元数据管理与数据质量管理平台



数据治理

数据治理和数据云



数据治理办法流程

- 复旦大学数据管理办法
- 复旦大学数据编码规范

数据资产平台

- 元数据管理平台
- 数据质量管理平台
- 数据安全管理平台

数据分析与服务平台

- 基础数据平台, 数据挖掘平台, 日志存储分析平台, 数据网关平台, 校园智慧信息大屏

提纲

- 背景——何为数据治理
- 要点——治理体系核心
- 实践——数据应用服务
- 探索——未来发展思考



教育行业的数据工作依然存在着诸多困难

- 部分院系、部门领导，对数据工作未能完全理解
- 从事数据维护一线人员，工作压力大、积极性低
- 信息化的技术层面上，定义不够明确、含义不清
- 缺乏既懂教育管理、又了解业务的信息技术人才
- 学校领导层面上，数据服务需求有待进一步明确
- 分清**之上而下**的规划与具体实施**两端并进**间关系

信息化部门在数据化管理中的角色

- 只有业务部门才能真正用好数据，做出有利于改进服务的分析
- 随着业务发展，对于数据的需求越来越多
- 技术部门不可能完全代替业务部门进行数据、业务分析，应该以提供工具、搭建基础平台为主，抛砖引玉、而不是越俎代庖
- 技术部门应大力推进数据化管理队伍建设，完善数据服务平台，以数据获取的有效性、可用性、实时性为考核指标

数据驱动业务发展的误区与可视化的误区

- 拥有大量的数据不一定就能驱动业务发展
 - 数据质量差、脏；数据与业务关联性不大；数据采集容易、使用难
- 准备数据分析可能占项目的80%时间和精力
 - 需要将数据集中，希望能及时更新和更改、希望能使用相同数据源，这使得创建数据存储不再那么简单
 - 建立了分析团队不一定就能找出问题所在，漂亮的报表展现不一定就能最好解决问题
- 数据可视化并不是数据分析，数据可视化更不是大数据全部
 - 可视化工具专注于报告、而不是分析，需要预先制作的中央数据库，需要背景阶段的准备和数据的加入，不能成为唯一组件
 - 端到端解决方案是处理准备数据的工具和算法，是创建数据可视化报告的前端



报表并不是决策支持

报表的不足：

- 报表个性化可能**不足**
- 报表数据可能**不够**
- 报表维度可能**不全**
- 报表对比可能**不准**

不要拿ERP的报表
忽悠领导！

决策支持 = 图形报表分析 + 实时数据 + 判断决策

- 基层主要关心业务过程，而高层、中层则关注不同的维度和指标，做不同深度、范围的决策分析
- 数据决策支持目的不仅仅是让高层做决策分析，同时也要让中层管理者做决策分析，并让高层看到中层是如何运作、管理、决策、分析

处理好数据管理、数据治理与数据质量管理的关系

数据治理是指在企业中，数据管理的相应职责和决策权的分配，是数据管理中最重要支柱之一，紧密把控数据质量；包括了数据发现和分析，目的是发现隐藏的数据质量问题

研究人员就数据治理达成了若干共识，即在数据治理过程中，必须回答以下三个问题：

1. 从全局出发，该做什么样的决策

数据治理制定的决策是关于数据质量的要求和测量、元数据管理、数据获取要求管理以及数据生命周期管理等一些数据管理的基本原则

2. 决策制定过程中，涉及哪些角色

指出数据治理过程中包含的角色大多分为数据管理人员、数据拥有者和数据委员会

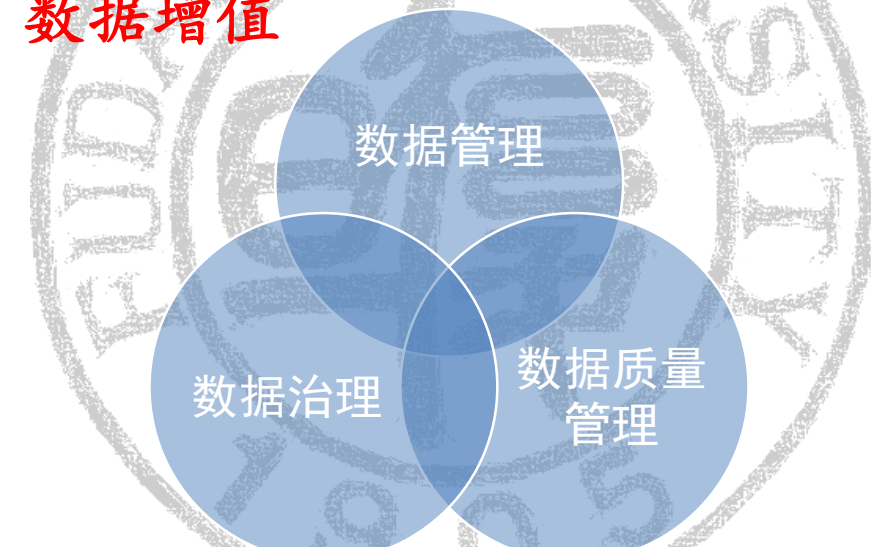
3. 这些角色是如何发挥作用的

数据治理涉及权力与职责的分配

数据治理旨在将数据的价值放大到最大化

数据质量是数据质量管理的结果，数据质量的最大化有利于数据价值达到最大化

三者的关系相辅相成，缺一不可，其共同目的都是使数据增值



不懂业务就不要谈数据分析

数据分析师、或想从事数据工作，一定要设法去深入了解业务知识

- 如果不懂业务，你连别人的问题都听不明白，连数据都不知道取哪一个
- 如果不懂业务，你都提不出一个好建议，又如何能让业务来依靠你
- 如果数据不结合具体业务知识、就是一堆数字，不会产生任何价值

核心：简化业务流程、规避技术复杂度，而非简化IT流程

**充分理解业务集成与数据整合的不同、难易程度

数据/数字是客观的，但对数据的解读则带有很强的主观意识：懂业务才能懂数据，在数据服务层面业务专家和技术专家同等重要

- 技术不可能完全代替业务，抛砖引玉、辅助决策，而非越俎代庖

什么是中台技术？

➤ 中台的来历

- 2015年马云带队考察了芬兰Supercell公司，不到200人的企业却能快速开发出好的游戏作品：原因是他们把公共的通用的组件做了积累，每次创作新产品时无需重复写这些基础组件，这样开发速度大大提高

➤ 中台的本质也是为了降低开发成本和提升效率

- 业务视角：业务能力的共享和复用
- 技术视角：软件、数据服务组件的复用

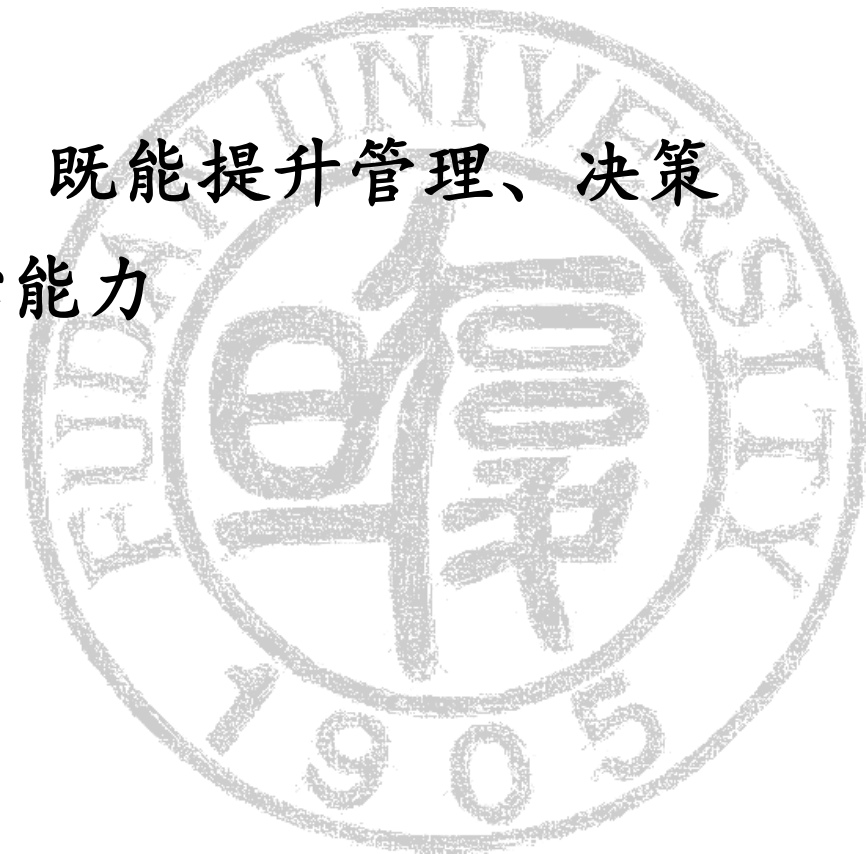
➤ 把相同的公共资源(组件)抽取出来、进行共享(复用)，由最初的竖井式(烟囱式)架构逐步演变成平台化结构

- 通过一缓冲层，把相关信息同步到平台，前端直接访问平台，中间层平台再和后台采用异步消息机制来减轻对业务系统的压力，从而解决业务系统(ERP)处理并发能力不足、或对资源共享支持不够的问题(教育：2C + 2B)
- 大中台适合范围：很多类同业务板块、面向2C端业务
- 面向2B的业务一般不一定非要建复杂、庞大的“中台”，不要一味追求时髦！



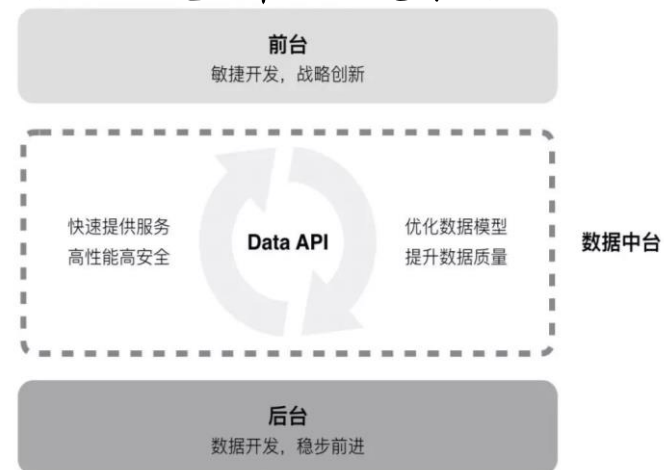
数据中台

- 数据中台是一种“把数据有效用起来”的机制，是一种战略选择和组织形式
- 数据来自于业务，通过不断循环迭代，实现数据可见、可用、可运营，反哺业务，让应用数据更加便捷
- 希望通过数据中台把数据变为一种服务能力，既能提升管理、决策水平，又能直接支撑业务，需要具备4个核心能力
 - 数据汇聚整合
 - 数据提纯加工
 - 数据服务可视化
 - 数据价值体现



数据中台不是大数据平台！

- 数据中台被誉为大数据下一站，由阿里兴起、核心思想是数据共享，并在2018年因“腾讯数据中台论”再度成为谈论焦点
 - 中台不是一个平台、也不是一个系统，是一种中间层；是为了弥补数据开发和应用开发之间、由于开发速度不匹配，出现的响应力跟不上的问题，为了让前台开发团队的开发速度不受后台数据开发影响
 - 数据中台以业务场景为颗粒度和思维模式建模，并将其封装成数据API服务，前台业务系统直接调取数据API服务产生业务价值，数据仓库搭建在数据API服务基础之上，做管理决策分析
 - 中台要解决的：效率问题(数据源常变)、协作问题(共享其它组已经开发的数据)、能力问题(专业独立的数据处理和维维护)



数据仓库、数据平台和数据中台的区别



数据中台可以小而美、面向场景而非面向技术，从小数据、小场景启动，演进式发展

数据创造的价值

行动早一点！

数据治理策略

创造的价值

规避的风险

先干再说！ 什么情况？ 原来如此...

管好数据！

时间



建立和完善数据化管理机制

- 校领导重视，学校层面主推数据业务整合
- 信息化部门建立专职数据化管理队伍，配备：
 - 数据分析/协调员：负责数据业务联系、反馈、跟踪和简单数据录入和分析服务
 - 数据库管理员：负责数据平台管理、质量控制和技术维护
- 数据手工录入过渡到推广数据托管平台(exchange)
- 数据生产利用需要吸纳需求迫切、最为活跃的用户，**倒逼**业务部门改变一个人数据中心等
- 借助院系二级管理的契机，把握院系自主决策的数据需求，通过**示范**效应，将部分试点院系吸引住，推动上层管理部门逐步迈向数据开放
- 多重手段推动数据化管理的可持续发展

数据管理需建立良性循环机制

- 所有信息化业务运行都是建立在相对精确的数据基础之上，数据质量问题不是仅仅通过IT部门的技术手段就能解决
- 数据只有在流转中才能体现其价值，有价值的数据才能引起使用者的关注，要使数据权威生产者**变被动为主动**地去维护数据，以保证数据有效性
- 通过在IT部门和业务部门间形成互助互利的数据交流模式，提高业务部门提供业务数据的积极性；IT部门通过对数据的收集和整理，反过来为学校 and 业务部门反馈数据的统计和分析结果，相互推动工作的发展，建立“个人-院系-学校”自下而上的数据监督与管理机制，形成内需拉动的信息化发展良性循环机制

加强数据治理顶层设计规划

➤ 领导抓数据治理工作

- 通过校情报告会、处级干部培训、下发文件、组织专题讨论等形式，强化各级管理部门信息化意识、数据治理意识
- 行政手段确保落实，综合办公室牵头进行相关工作协调，督察督办

➤ 业务部门提出数据治理模型

- 管理部门就管理领域内数据建立分析模型，帮助部门内部工作
- 多部门联合讨论建立交叉数据分析模型，为学校发展出谋划策

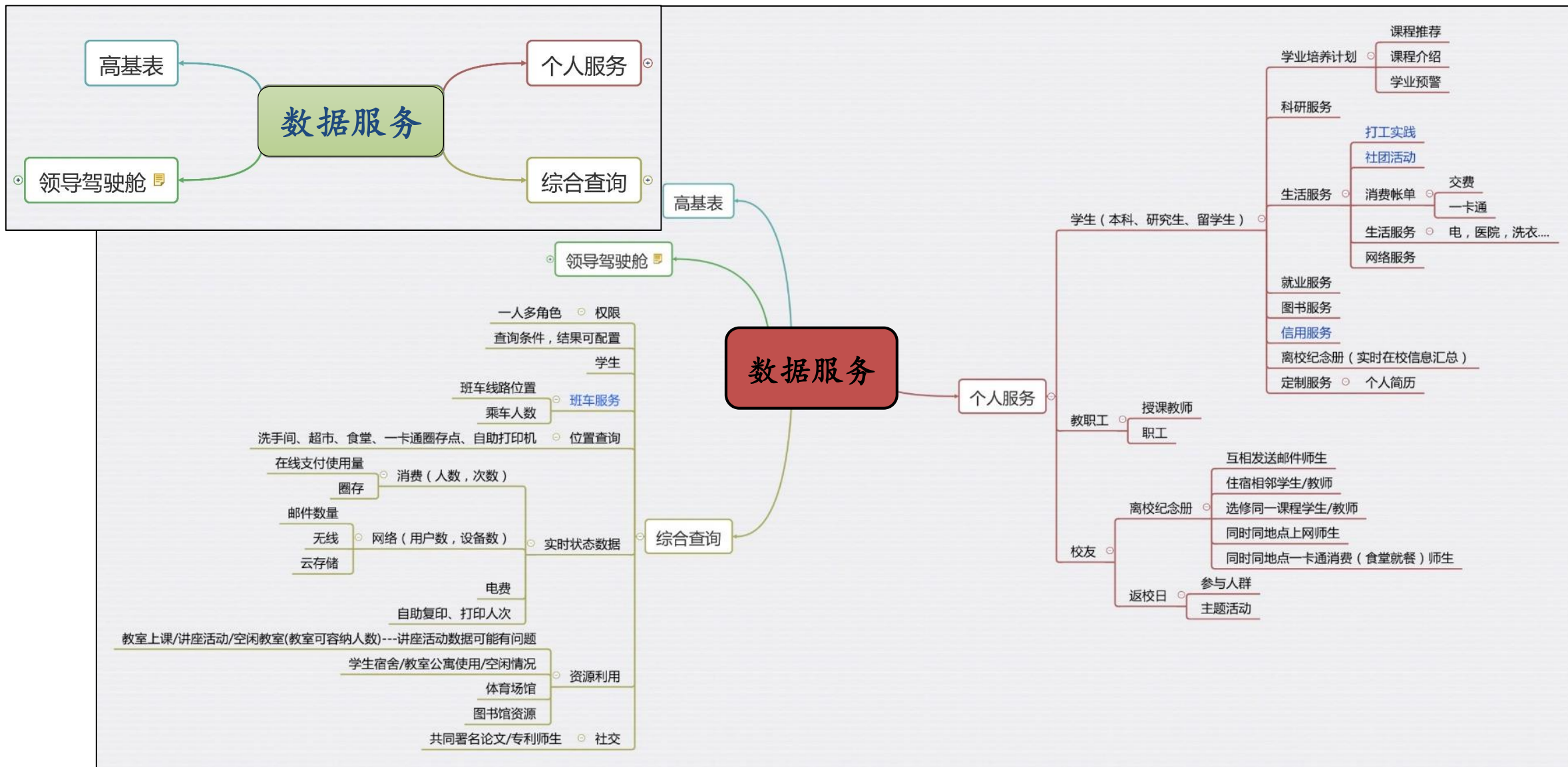
➤ 技术部门合作实施数据治理规划

- 建立统一的数据中心，搭建数据共享平台，完善数据共享机制
- 建立面向不同服务对象的数据管理与服务平台，提供数据可视化工具，深化数据挖掘和利用

决策需求主导数据服务方向

- 重视数据分析与挖掘对学校管理改革的支撑，有的放矢、真正为学校决策支持提供数据服务
- 信息化部门应主动出击、抛砖引玉，为学校、业务部门等不同层次的用户提供定制的主题数据分析，形成常态化工作，并逐步使行为主体从信息化部门转变为管理职能部门
- 让学校和业务部门切实体会到信息化带来的变化与好处，信息化部门也因此会成为学校管理和教育教学改革的真正推手

数据应用服务梳理与分类



有效应对智慧的需求

数据：从被动收集 → 到主动产生

➤ 目前的智能驾驶技术，主要集中于计算机视觉领域(智能)

➤ 复杂车辆自动行驶过程中，涉及车道线识别、行人识别、车辆识别、物体识别、标识识别等多方面技术

➤ 还需对复杂气候条件下道路特殊情况做出应对，如下雨积水反光、雪天遮挡车道线等

➤ 实现无人驾驶的“聪明”汽车，需要知道相关信息(智慧)

➤ 我在哪里？

➤ 我周围有什么？

➤ 接下来会发生什么？

➤ 我要怎么做？



复旦的探索

复旦大学这些年来一直在做这方面的探索，特别是2015年9月份在国内高校中率先推出了一站式“网上办事服务大厅eHall”，通过业务梳理、流程再造、数据集中、服务集成，结合互联网+思维、大数据/微服务等信息化新技术和智能终端/自助打印等新形式，推出了近百项师生网上一站式服务(这两年超过了200项)，实实在在迈出了坚实的一步，得到了广大师生的好评和领导、同行的关注，通过开展这项工作，自身也深有体会和感触、并得到了有效提高

面向用户服务的门户发展方向

由信息发布门户向办事服务大厅/一站式转变
—— 协同工作、提升核心竞争力



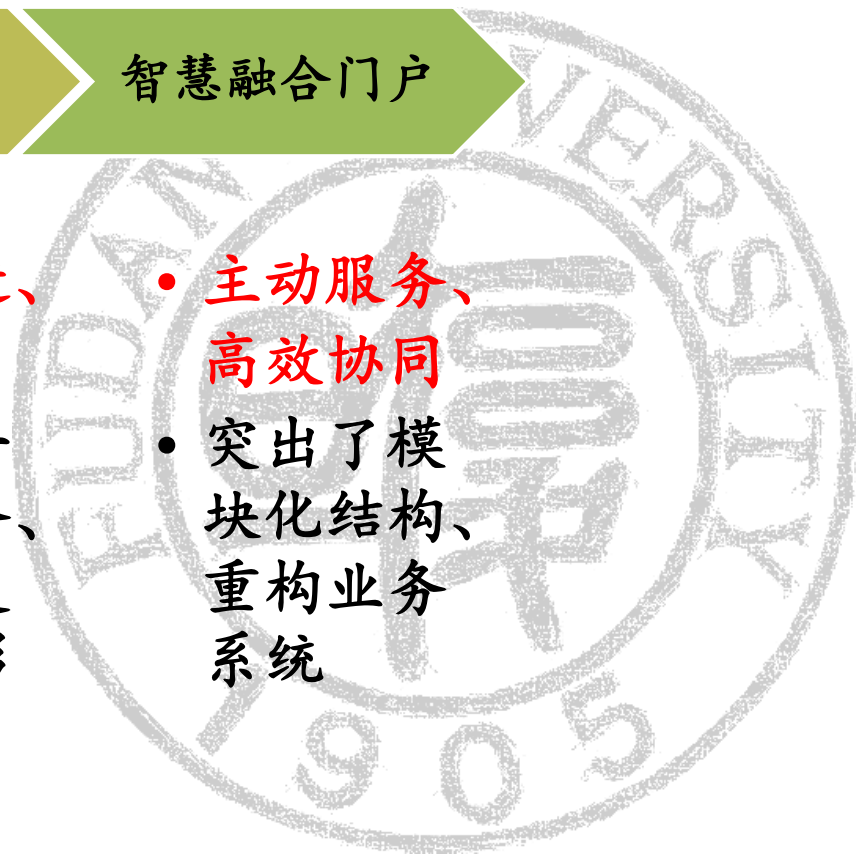
- 新闻抓取、信息聚合
- 提供了单向的信息聚合与发布服务

- 应用集合、单点登录
- 集成了个人信息中心、部门数据展示

- 服务分类、应用集成
- 初现了迎新、离校等跨部门服务特征

- 流程再造、服务碎片
- 展现了一站式服务、平台化支持的雏形

- 主动服务、高效协同
- 突出了模块化结构、重构业务系统



实施路线



2003

URP应用集成门户：服务与管理合并建设，
用户体验差



2009

Portal综合服务门户：服务与管理逐渐
分离，服务与服务之间关联较弱



2015

eHall一站式服务：采用平台化
架构建设，横向贯通服务流程



在已有工作基础上，扩大服务广度和深度，提升服务水平
挖掘数据价值，为学校发展赋能

2019

深化建设



“一站式”网上办事服务大厅

- 体现“**服务优先**”理念，注重关心与师生利益密切相关事务
- 精简和优化流程，将**IT技术手段转化为良好的用户服务体验**
- 面向全校师生服务，分多批上线服务开放运行、征求意见



大幅减少师生到处寻找、部处日常重复工作问题，提升服务效率



统一流程中心，可看到待办任务和流程追踪

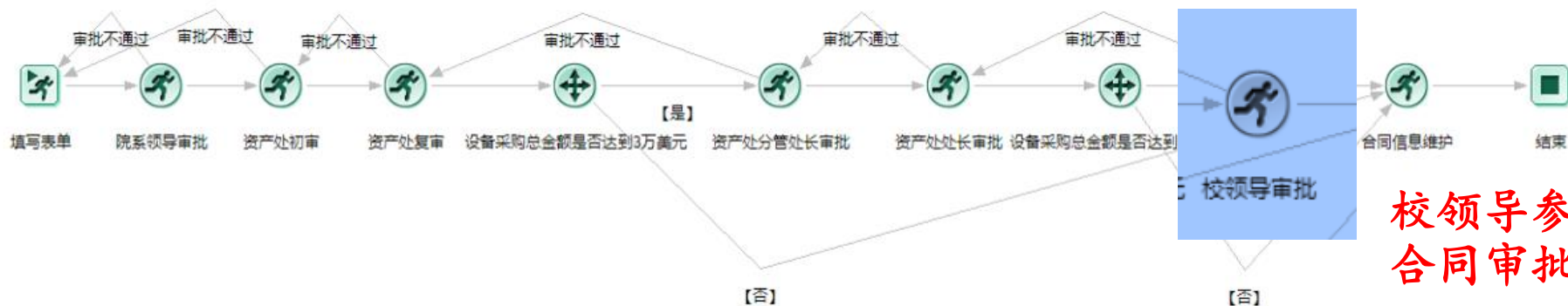
按角色分为学生办事和教师办事两块，支持游客访问

<http://ehall.fudan.edu.cn>

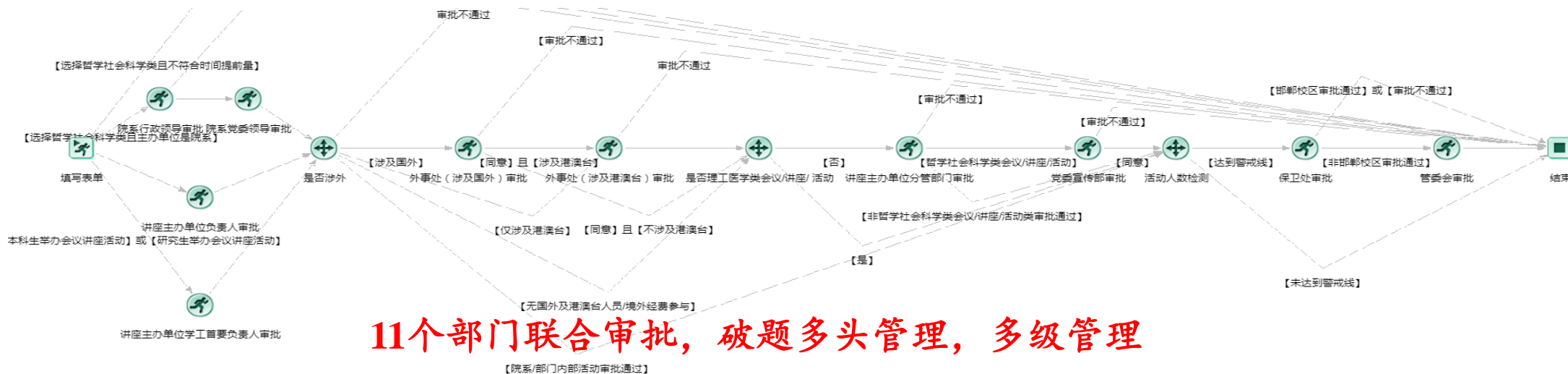
推荐服务、最新注册服务、热门服务

做好服务“一站式”

➤ 越来越多线下流转的跨部门流程开始线上流转



校领导参与到外贸合同审批业务中



11个部门联合审批，破题多头管理，多级管理

- 上线服务**206项**
- 提交申请**近10万项**
- 职工自助打印材料**3.4万份**
- 本科生自助打印材料**13万份**
- 研究生自助打印材料**7.1万份**

eHall线上线下载融合O2O



復旦大學

出国、赴港澳任务批件

复旦外出[2019]2488号

出访团组名称	何欣怡等3人赴德国交流学习	负责人	何欣怡	
		出访日期	2019年08月31日-2020年01月30日	
		出访人数	3	
出访任务	交流学习			
出访及途经地点	出访国家或地区	德国(科隆)		
	途经国家或地区	停留153天		
邀请接待单位名称	科隆大学			
费用来源	一切费用个人自理			
姓名	性别	出生日	工作单位	职务或职称
何欣怡	女	1990.08.08	复旦大学管理学院	本科生
何欣怡	男	1990.08.08	复旦大学管理学院	本科生
何欣怡	女	1990.08.08	复旦大学管理学院	本科生
以下空白				
政审号	沪复委审字[2019]2600号			
附注	此件三个月内有效,出访时间含离境当日。			

抄送单位: 市政府外办、市财政局、外汇管理上海分局

二〇一九年七月廿四日



线下自助服务

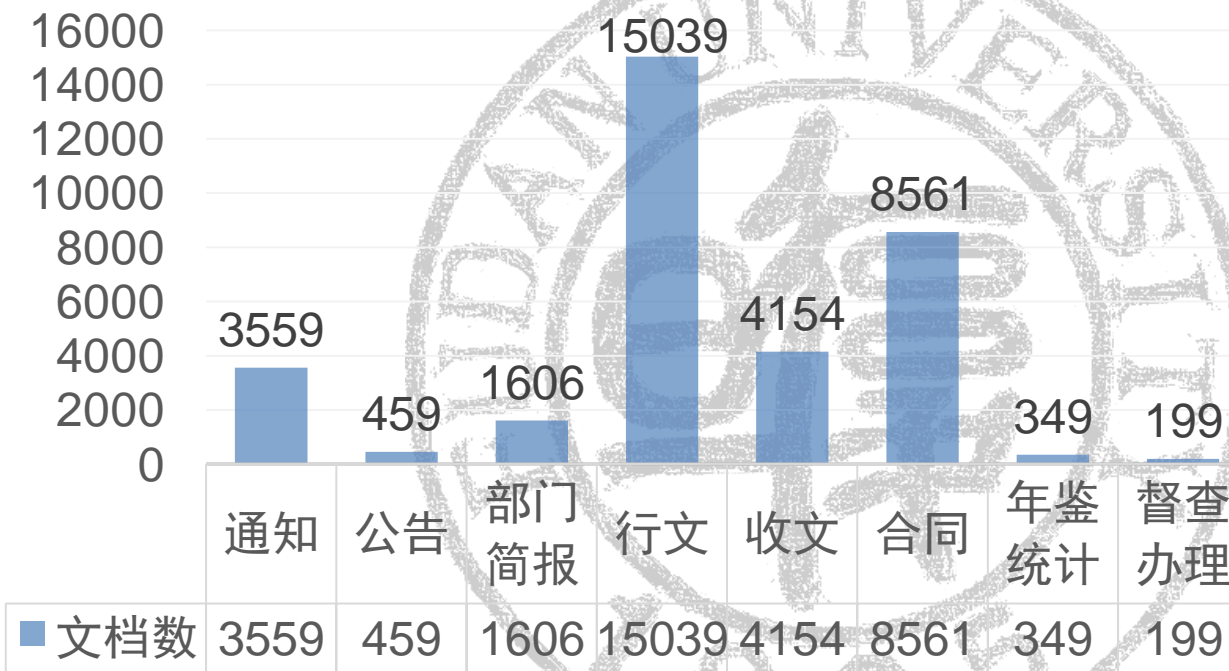
典型流程



数据重构、流程重组、服务再造
线上服务改革

校务治理---办公自动化OA

复旦大学校园电子公务系统



复旦大学新版校园电子公务系统从2016年1月开始正式启用

校务/服务治理

各部门深度协同

64

业务系统

全面支持教学、
科研和校务管理

200+

师生服务

提供
全生命周期服务

97

支持单位

覆盖21家职能部处
18家附属单位
58个教学院系

5.2万

用户数

覆盖
全体师生

实现资产报废、调拨、变更等线上业务流程，优化资产管理流程

和资产处内贸合同系统对接，实现合同信息提取

和财务系统实现数据交换，做到帐帐相符

为解决科研项目“报销繁”问题提供技术指导

大资产管理



校务治理---大人事管理

进校前

在校期间

离校后

管理岗招聘:

2019年上半年党政管理岗位招聘,共677人次报名,297人进入笔试环节

派遣(租赁)招聘业务上线,已面向全校机关部处和院系进行了培训

人事处网站

人才招聘系统

人才招聘网

光华青年学者论坛

人员进校申请系统

面向人事处用户

人事管理系统

教职工信息管理

科级干部任免

公派出国信息管理

校内调动管理

人事档案管理

博士后考核管理

数据采集系统

Exchange平台

面向教师用户

eHall网上办事服务大厅

个人数据中心
教育部信息采集

岗位考核
校外兼职审批登记

校外同行评议
机关测评

博士后管理
博士后出站申请

自助打印服务

其它服务

教职工在职证明

职称证明

在线缴纳党费

中英文在职证明

博士后在站证明

在线缴纳幼儿园费用

人事管理系统

离校管理

退休管理

返聘管理

延聘管理

死亡管理

协同了包括人事处、组织部、统战部在内的多部门构建**校级大人事系统**,实现了数据统一和教职工全生命周期跨业务域的服务和管理

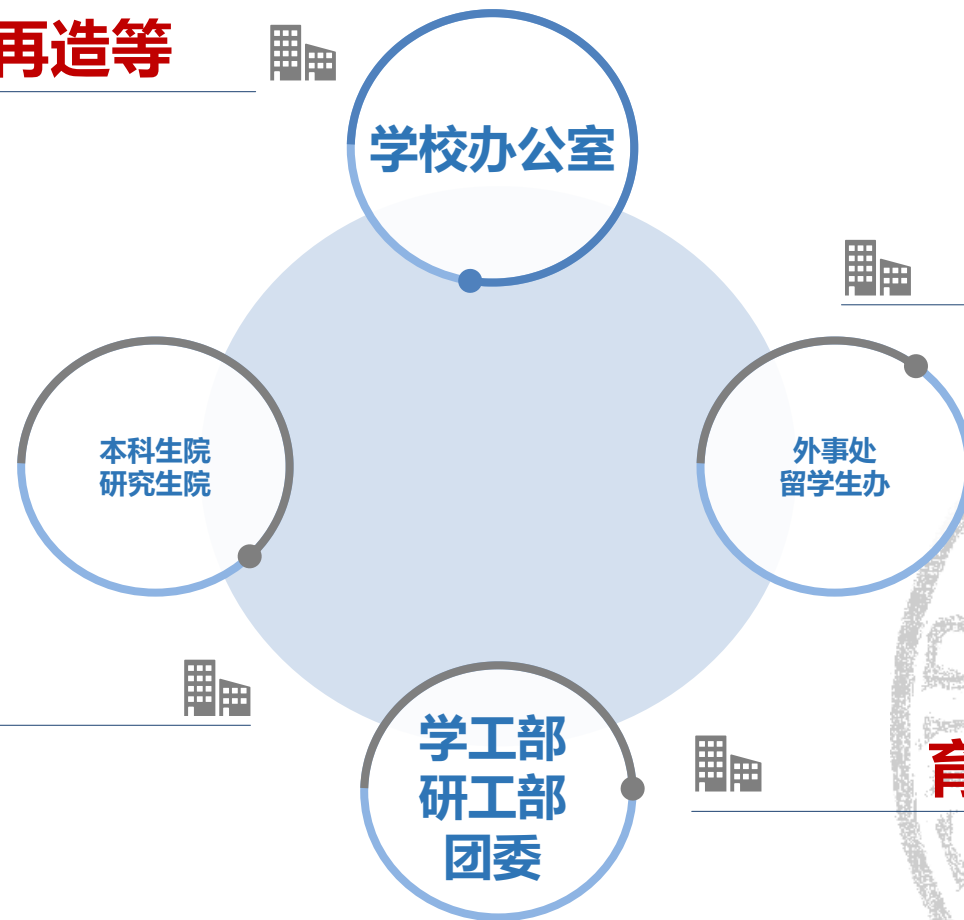
校务治理---大人才培养

牵头服务梳理、流程再造等

- 一网通办
- 校园公务电子系统 (OA)
- 智慧校园数据云
- 个人数据 “一表通”
-

教育教学改革

- 本科22条
- 本研贯通
- 智慧教室
- 本科教学实验中心
-



国际化

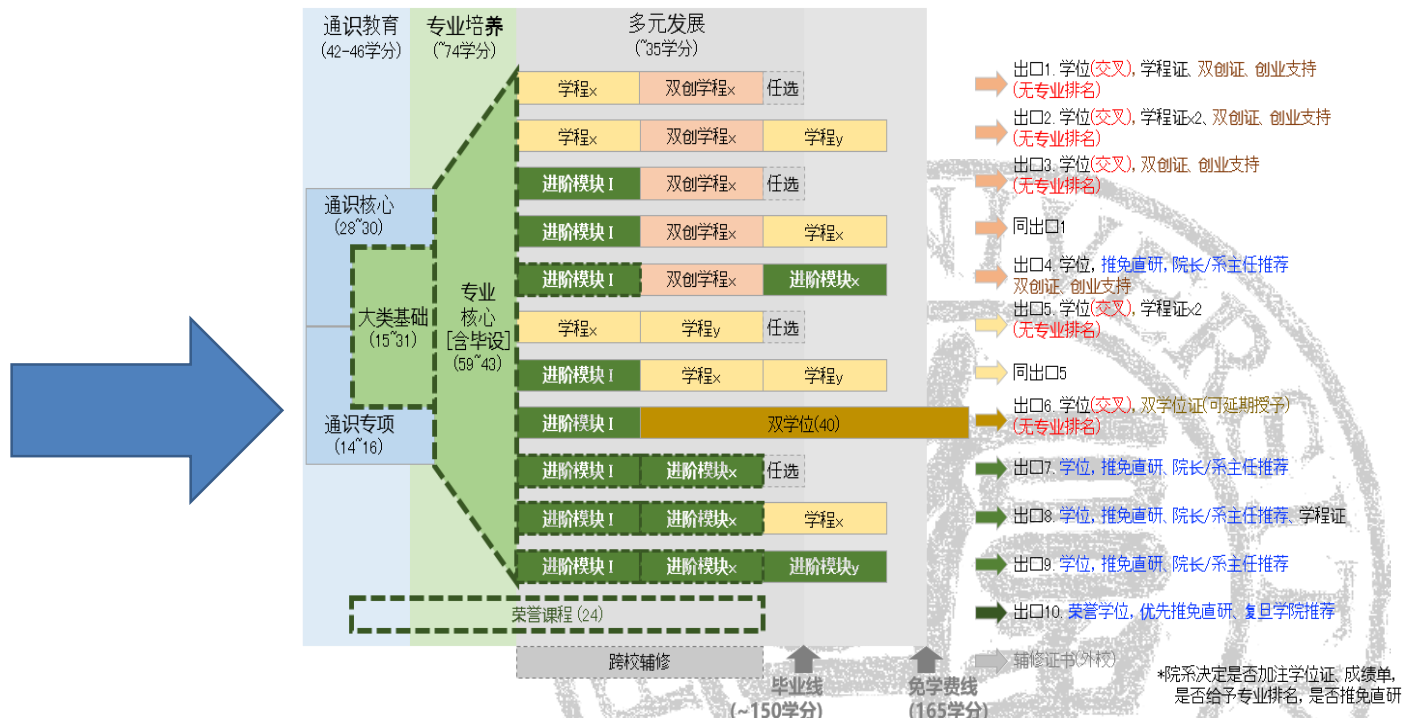
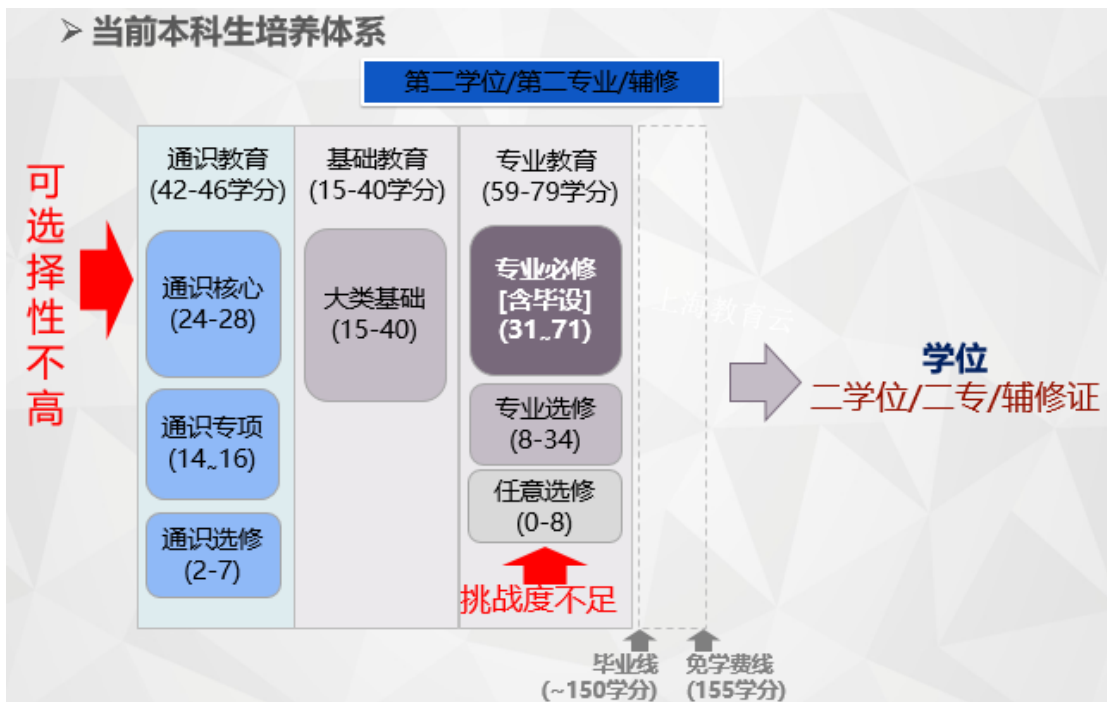
- 项目交流
- 留学生全生命周期管理
- 系统英文版建设
-

育人与社会实践

- 第二课堂育人
- 社会实践平台
- 思政工作信息化平台
-

校务治理---本科生教学改革

学生的培养路径：2个选择→13个选择
学生的可选学分：42个学分→109个学分



根据学校改革政策，对人才培养平台中教务系统的培养方案、课程管理、教学运行、毕业审核、导师管理等进行升级和整合

校务治理---灵活适应培养方案调整的选课系统

- 学分制选课：**全部课程都可以纳入选课
- 交互式选课：**学生实时掌握课程资源数量
- 多批次选课：**多个选课批次可以同时开放
- 国际化选课：**留学生、暑期国际课程英文版选课
- 高性能选课：**全部采用虚拟化部署、高并发



高并发

- 本科生每学期3300门课程，选课平均14万人次
- 本科生选课系统最高在线人数8500人/秒



本研打通

- 2017年9月，系统支持本研互选。
- 试点院系的大四本科生可以提前选修部分研究生专业课；
- 研一新生可以补修本科生课程



业务解耦

- 利用新媒体技术，结合微信公众号，将课程成绩查询从选课剥离，提升了学生用户体验，降低了选课系统负载

校务治理---智慧教室电子班牌



门禁管理
使用分析
签到管理

空间使用状态
信息发布
预约管理
服务跟踪



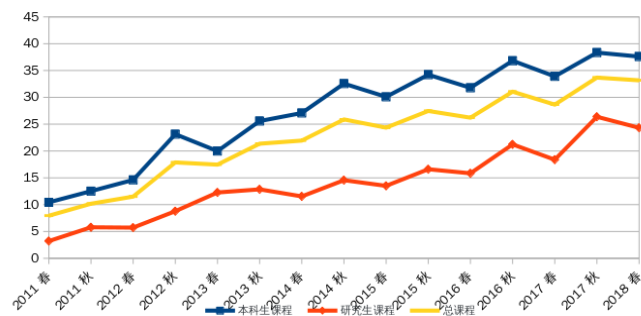
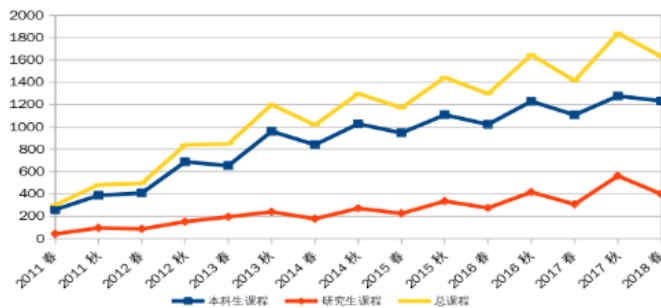
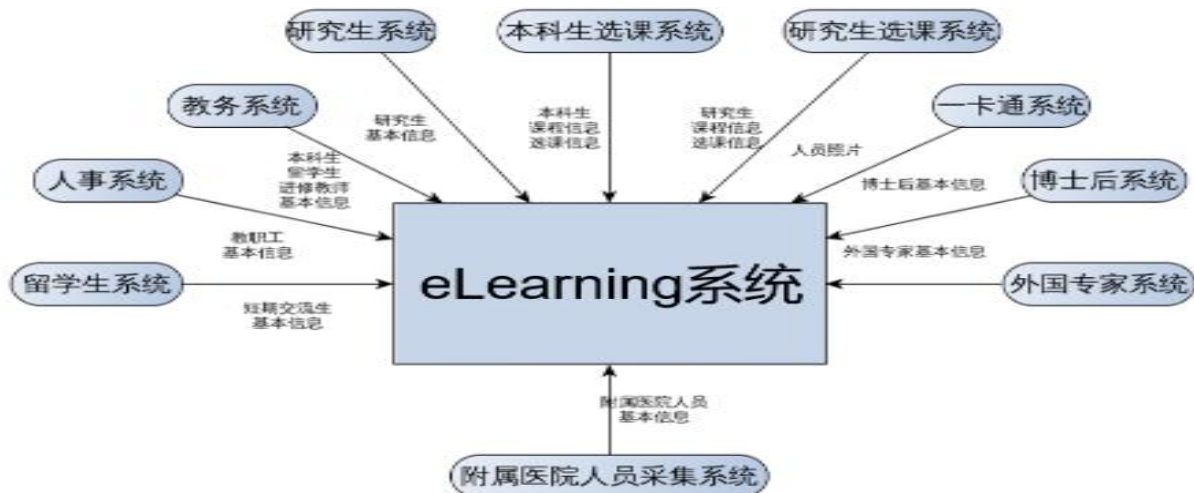
- 全面提升智慧教室利用率，协助学校进行运营管理
- 以服务为导向，兼顾智慧教室服务效率及服务品质
- 多元签到，刷卡
- 通过智能终端，实现校园服务整合

学校系统



UIS
教务系统
一卡通管理系统
eHall平台

校务治理---eLearning在线学习系统



eLearning平台于2010年基于开源软件Sakai建设，与本科生教务、本科生选课、研究生教学管理、选课、人事管理、一卡通、外国专家等多个系统打通，实现**教学和教务管理一体化**；为教学提供网络辅助技术手段



使用的教学班数量和百分比逐年上升

- 累计服务**>15000**个教学班
- 累计使用用户**9.4万人**
- 累计访问人次**964.8万人次**

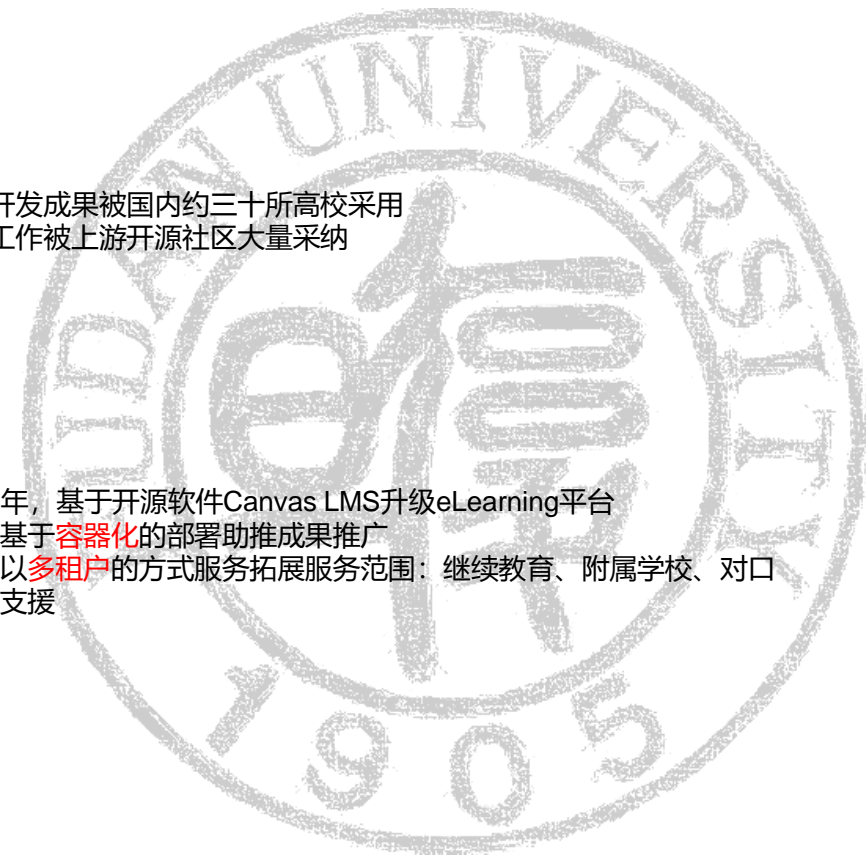


相关开发成果被国内约三十所高校采用
相关工作被上游开源社区大量采纳



2018年，基于开源软件Canvas LMS升级eLearning平台

- 基于**容器化**的部署助推成果推广
- 以**多租户**的方式服务拓展服务范围：继续教育、附属学校、对口支援

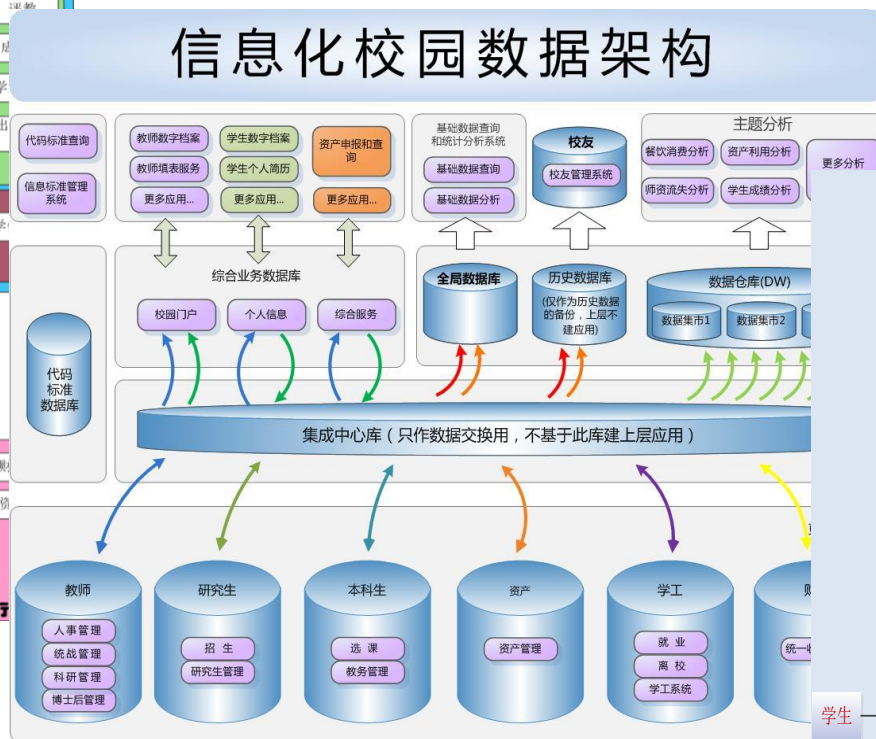


服务、数据、流程的梳理

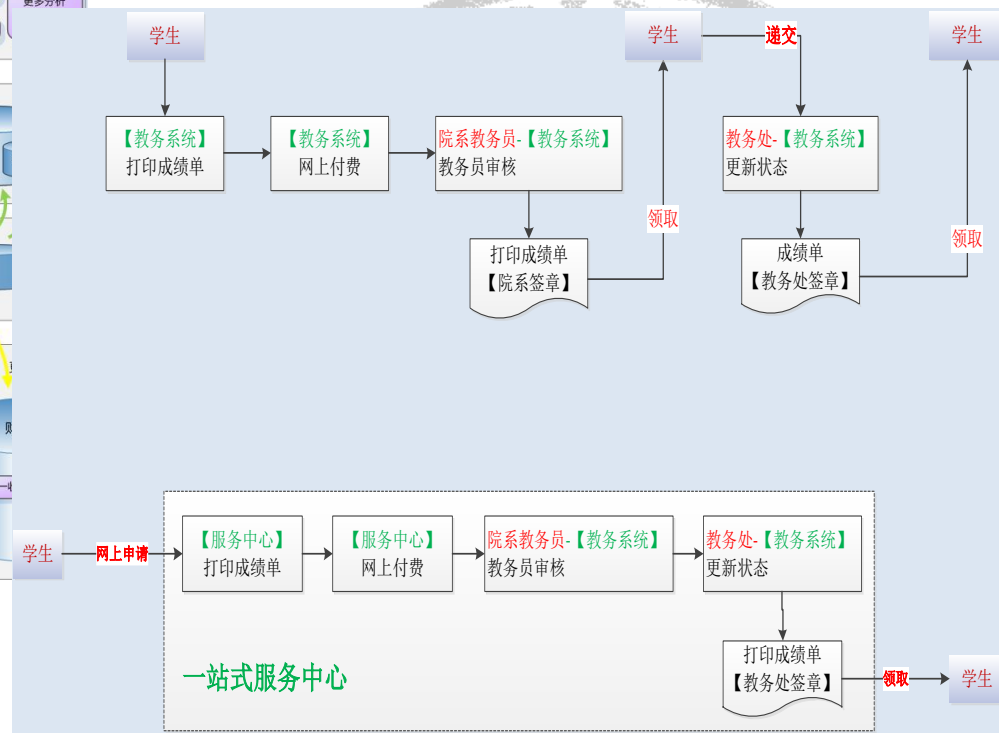
➤ 服务梳理



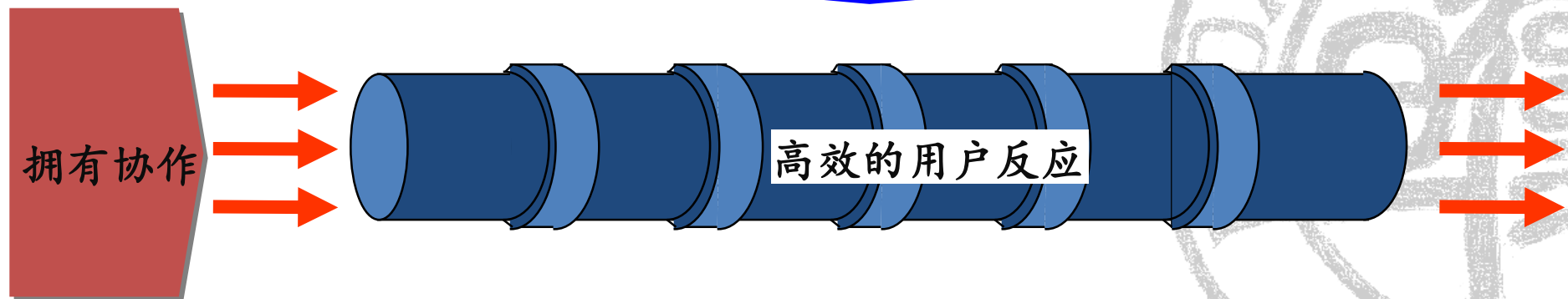
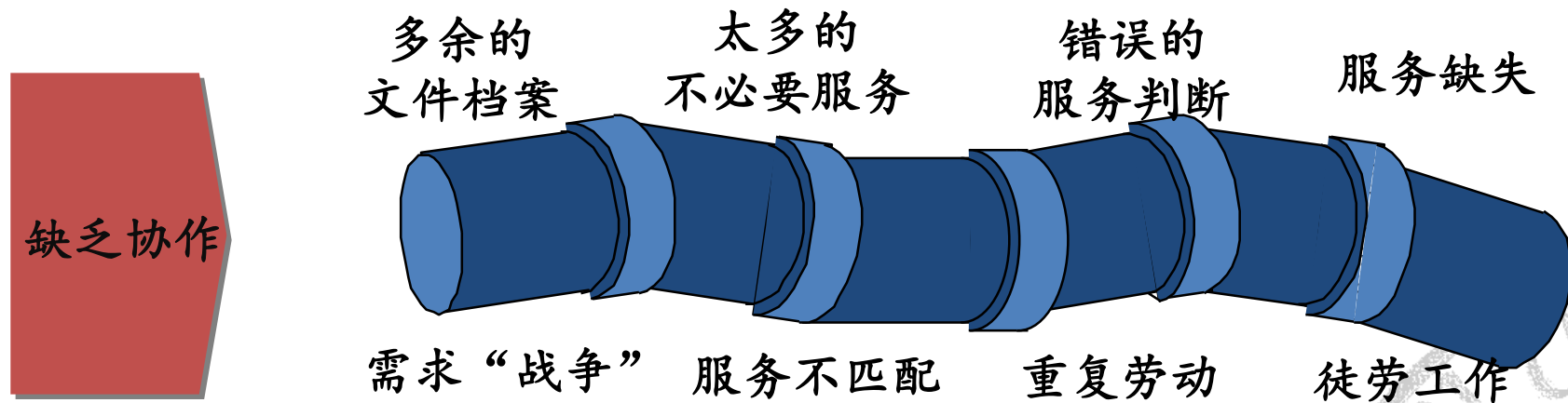
➤ 数据梳理



➤ 流程梳理



高效协同工作的条件



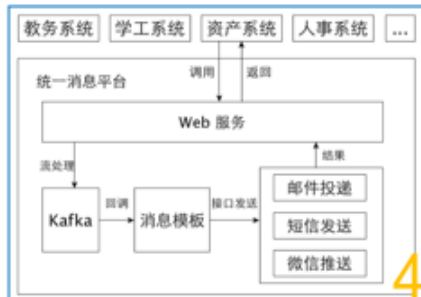
本人以为 → 以人为本

技术支撑平台



统一身份平台

- 解决不同应用系统间用户名和口令不统一问题
- 界面支持**中文和英文**两种语言；
- 激活状态的账号总数**>10万**，接入的正式业务系统有**95个**



一表通平台

- 信息办自主开发的表单填写系统；为eHall提供填写入口
- 配置简单，开发速度快，可为特殊需求定制开发新功能
- 统一界面、提升系统一致性

workflow平台

- 工作流平台提供统一的表单、流程、视图的定制开发
- 可跟踪流程进度
- 每个流程可通过邮件、短信、微信提醒，及时获得流程进展

统一消息平台

- 统一消息平台支持**邮件、短信、微信**等多种通知方式
- 短信平台为全校各部门提供了统一的短信发送以及查看的平台
- 采用通用开放接口为其他应用系统提供消息发送的服务支撑



统一收费平台

- 提供**7*24小时**在线支付服务
- 网上自助缴费**>344万笔**、金额**>31亿**
- 能满足英文界面的**国际卡跨国支付**要求，方便了国际学术交流

自助服务平台

- 为学生和教师提供自助服务；其平均打印业务办理时间为1分钟
- 四校区共计**27台**，提供**15项**自助服务
- 共打印材料：教职工**3万份**，本科生**12万份**，研究生**6.4万份**

元数据管理与数据质量管理实施效果

复旦大学信息化校园系统

共享数据库

信息编码规范说明书

文件编号 (WISCOM-W51A-SC02-K03_DB-02)

版本号 (V1.5)

发布时间 (2003.10.16)

编写人 (徐东生)

读取器配置

读取器源 图标和系列

类型 暂存数据

源 SJFWDB.USR_JZG_T_JZG_JBXX

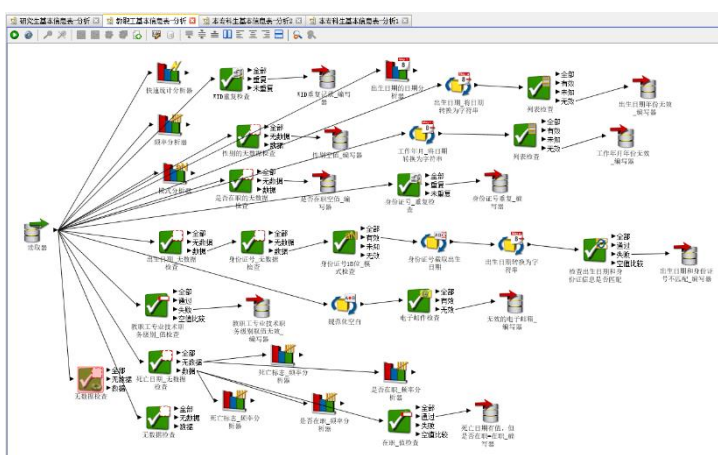
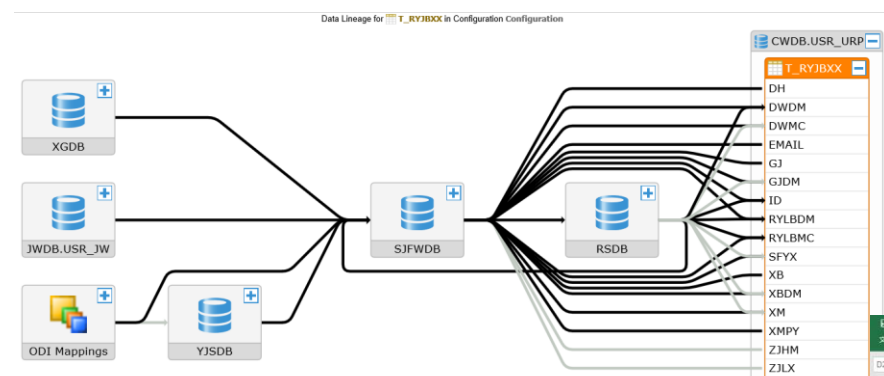
WID (string)
GH (string)
XM (string)
XMM (string)

Compare Target: SJFWDB - 2018-11-16 12:03:29 with Source: SJFWDB - 2018-10-29 15:49:57

Changed objects	Source (2018-10-29 15:49:57)	Target (2018-11-16 12:03:29)
Database	sjfwdb2	sjfwdb1
Schemas	USR_JXJY	USR_JXJY
Tables		CS_0STUDENTPLAN
Schema	USR_YJSY	USR_YJSY
Tables		CS_0ASA3D9ND369J5HLPT7C9P32764U

- ✓ 共享库信息编码规范
- ✓ 数据影响力溯源
- ✓ 数据结构监控
- ✓ 数据结构版本管理
- ✓ 数据质量审计
- ✓ 数据质量评估报告

- PRO_重要数据库
- Configuration
- AmodeI
- CWDB
- DZBP
- JWDB
- JXJY_YD
- PJXT
- RSDB
- SHDB
- SJFWDB
- 2018-11-16 12:03:29
- 2018-10-29 15:49:57
- 2018-10-18 09:03:27
- TYDB
- XGDB
- YJSDB
- ZHDB
- ODI Mappings
- 2018-12-03 15:49:40
- 2018-11-28 16:48:42
- 2018-11-23 16:33:45
- 2018-11-23 15:53:48
- 2018-11-23 14:34:29
- 2018-10-18 11:58:04



参数名	类型	说明	来源	空值率	其他	数据导出	备注
WID	CHAR(40)	系统唯一编号	XGDB.USR_JG.T.XSBJXX.XSBB.WID & XGDB.USR_JG2014.T.BKJ.BKX.WID & JWDB.USR_JW.C.STUDENTS.CODE	0			
XH	VARCHAR(20)	学号	XGDB.USR_JG.T.XSBJXX.XSBB.XH & XGDB.USR_JG2014.T.BKJ.BKX.XH & JWDB.USR_JW.C.STUDENTS.CODE	0	学号有四种模式，长度分别为7,10,11,12		
XM	VARCHAR(20)	姓名	JWDB.USR_JW.C.STUDENTS.NAME	0	姓名有四种模式，进一步进行模式本身正则表达式分析，发现部分姓名包含非汉字字符		姓名属性包含无效字符!
XMPY	VARCHAR(120)	姓名拼音	JWDB.USR_JW.C.STUDENTS.ENG_NAME	0	姓名拼音包含非汉字字符		姓名拼音包含无效字符!
ZYM	VARCHAR(20)	曾用名	N/A	99.70%	曾用名取值"上"1609条，"工"1009条，取值为多个空格的有20条		
YW	VARCHAR(120)	英文名	N/A	38%	识别出476条具有26个英文字母及空格之外的字符的记录		
XBDM	VARCHAR(21)	性别代码	JWDB.USR_JW.GB.GENDERS.CODE	0	值非"男"的有679条，0.7%		
XB	VARCHAR(20)	性别	JWDB.USR_JW.GB.GENDERS.NAME	0	值非"男"的有679条，0.7%		
CSRD	DATE	出生日期	JWDB.USR_JW.C.STD_PEOPLE.BIRTHDAY	2.30%	出生日期取值"上"1609条，"工"1009条		出生日期包含无效!
MZDM	VARCHAR(2)	民族代码	JWDB.USR_JW.GB.NATIONS.CODE	15%			
MZ	VARCHAR(2)	民族	JWDB.USR_JW.GB.NATIONS.NAME	15%			
ZZMMDM	VARCHAR(2)	政治面貌代码	XGDB.USR_JG2014.T.BKJ.BKX.ZZMMDM	85.10%			
ZZMM	VARCHAR(45)	政治面貌	XGDB.USR_JG.T.ZN8Z.ZZMM.ZZMMCMC - SJFWDB.USR_ZN8Z.ZN8Z.ZZMM.ZZMMCMC	85.10%			
SPZH	VARCHAR(20)	身份证号	JWDB.USR_JW.C.STD_PEOPLE.ICARD	17.6%	结合国家地区、学生类型、学院的信息检查应该具有18位身份证号		身份证号包含无效!
BH	VARCHAR(26)	班号	XGDB.USR_JG.T.BKJ.BJ.BKMC	84.20%	确定空值的身份证号4条		
SYDM	VARCHAR(26)	生源地代码	JWDB.USR_JW.GB.DIVISIONS.CODE	11.00%	身份证号应该有18位15位的国家地区		身份证号包含无效!
SYD	VARCHAR(30)	生源地	JWDB.USR_JW.GB.DIVISIONS.NAME	11.00%	身份证号应该有18位15位的国家地区		身份证号包含无效!
XQDM	VARCHAR(20)	校区代码	JWDB.USR_JW.C.CAMPUS.CODE	28%	身份证号应该有18位15位的国家地区		
XQ	VARCHAR(20)	校区	JWDB.USR_JW.C.CAMPUS.NAME	28%	身份证号应该有18位15位的国家地区		
HSSBK	VARCHAR(26)	宿舍楼号	N/A	78.50%	宿舍楼号有多种模式，比如"2017楼中文"，"18楼管理楼"，"2017楼数字1楼"，"理科2楼"等，建议对数据格式进行统一		宿舍楼号包含无效!
XSLB	VARCHAR(10)	宿舍类型代码	JWDB.USR_JW.XB.STD_TYPES.CODE	0			
XSLBMC	VARCHAR(10)	宿舍类型名称	JWDB.USR_JW.XB.STD_TYPES.NAME	0			

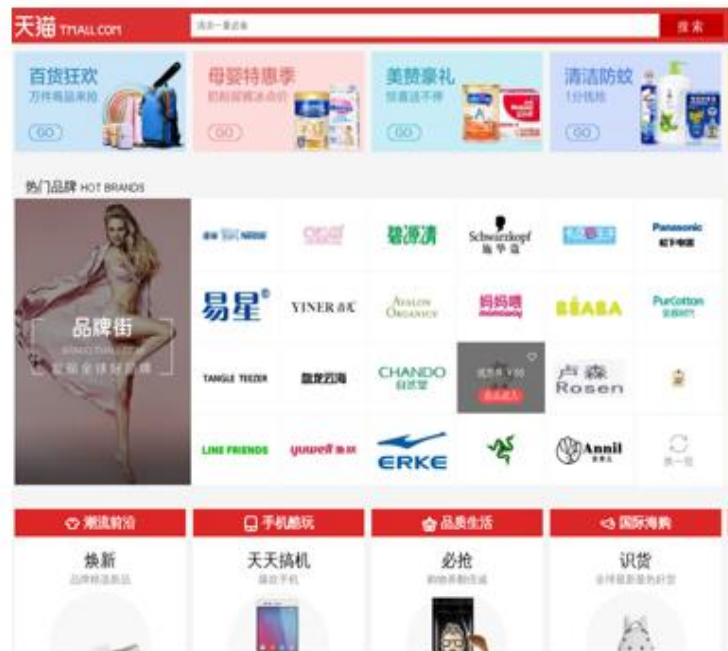
来自于复旦大学信息办张凯副主任

传统Portal方式

- 类似于**传统商场**，提供各个服务(品牌店)入口
- 各类业务(购物活动)在具体业务系统(品牌店)中完成
- 用户体验不一致
- 用户无法方便反馈服务质量
- 业务系统故障影响服务



工作模式转变： 让办事像网购



eHall模式

- 类似于**网上购物**，几乎所有的流程都在同一个入口完成
- 业务系统只关心抽出的数据(收订单、发货)
- 用户体验一致
- 用户可以直接反馈服务体验
- 业务系统故障不影响服务



一站式服务 One-stop Service – 转换视角

➤ 本人以为

- 部门视角(organizational perspective)
- 供给导向(supply-based)
- 管理导向(management-oriented)
- 把人当机器(human as machine)

➤ 以人为本

- 用户视角(customer perspective)
- 需求导向(need-based)
- 服务导向(service-oriented)
- 让机器像人(user-friendly)

➤ 让人别跑

➤ 把舒服留给自己

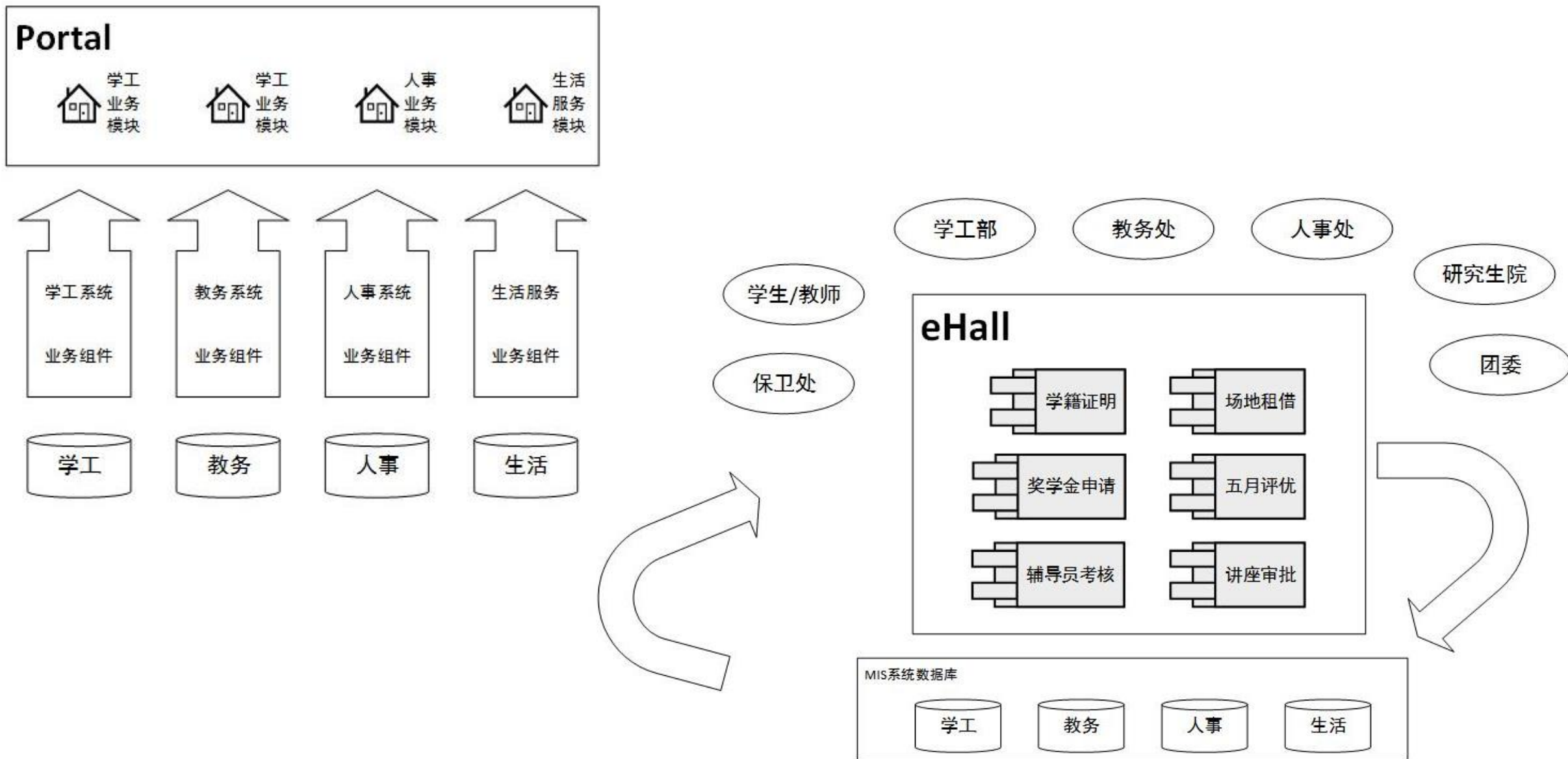
➤ 让数据跑

➤ 把方便带给用户

适应新模式的需求

- “一站式”服务平台突出的是服务体验，考验的是整体管理水平和平台化整合能力；面向师生用户服务的改造，不只需要考虑流程审批，更离不开数据的支撑
- 网上办事大厅的关键是面向用户的服务与管理分离，突出大平台、轻应用、微服务的特点，**数量**---并非一站式服务核心要素、**翻转**---才是发生革命性变化希望；数据打通比业务贯通更简洁、清晰
- 利用管理体制改革，推进业务流程整合、实现综合治理改造；依托数据集成共享和“碎片化”服务，抽取共性、形成基础服务
 - 强调信息资源在信息化建设中的核心地位和重要性
 - 以数据集中、信息整合、业务贯通和统一服务为核心理念

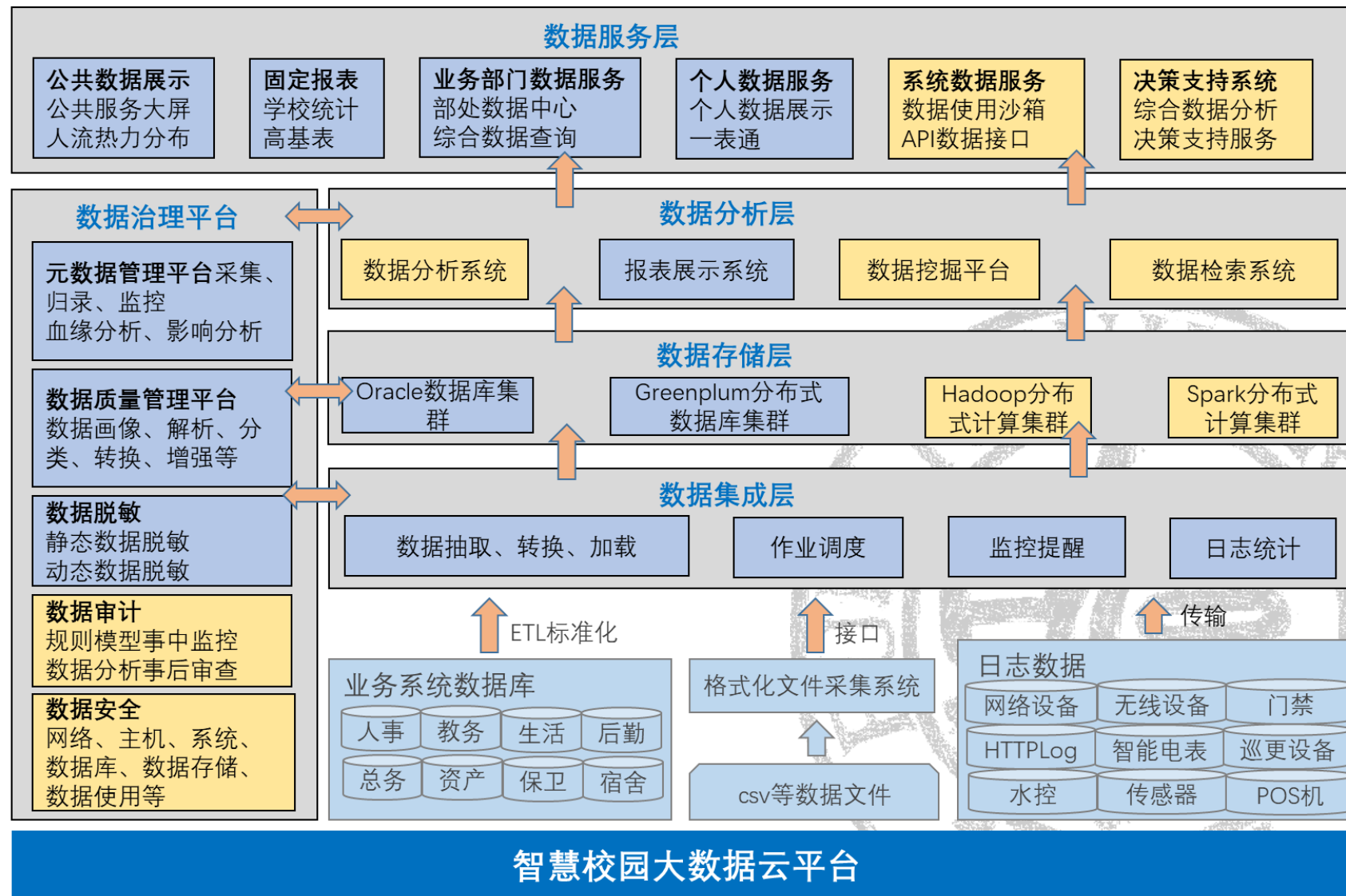
数据和业务流程处理方式对比



逐步做全数据“一张表”

- 数据资源是当代社会的核心战略资源
- 制定数据采集、归集、整合、共享、开放、应用以及质量和安全管理等标准规范
- 建设数据云平台，管理学校数据资产

数据标准与安全规范



教育数据管理办法

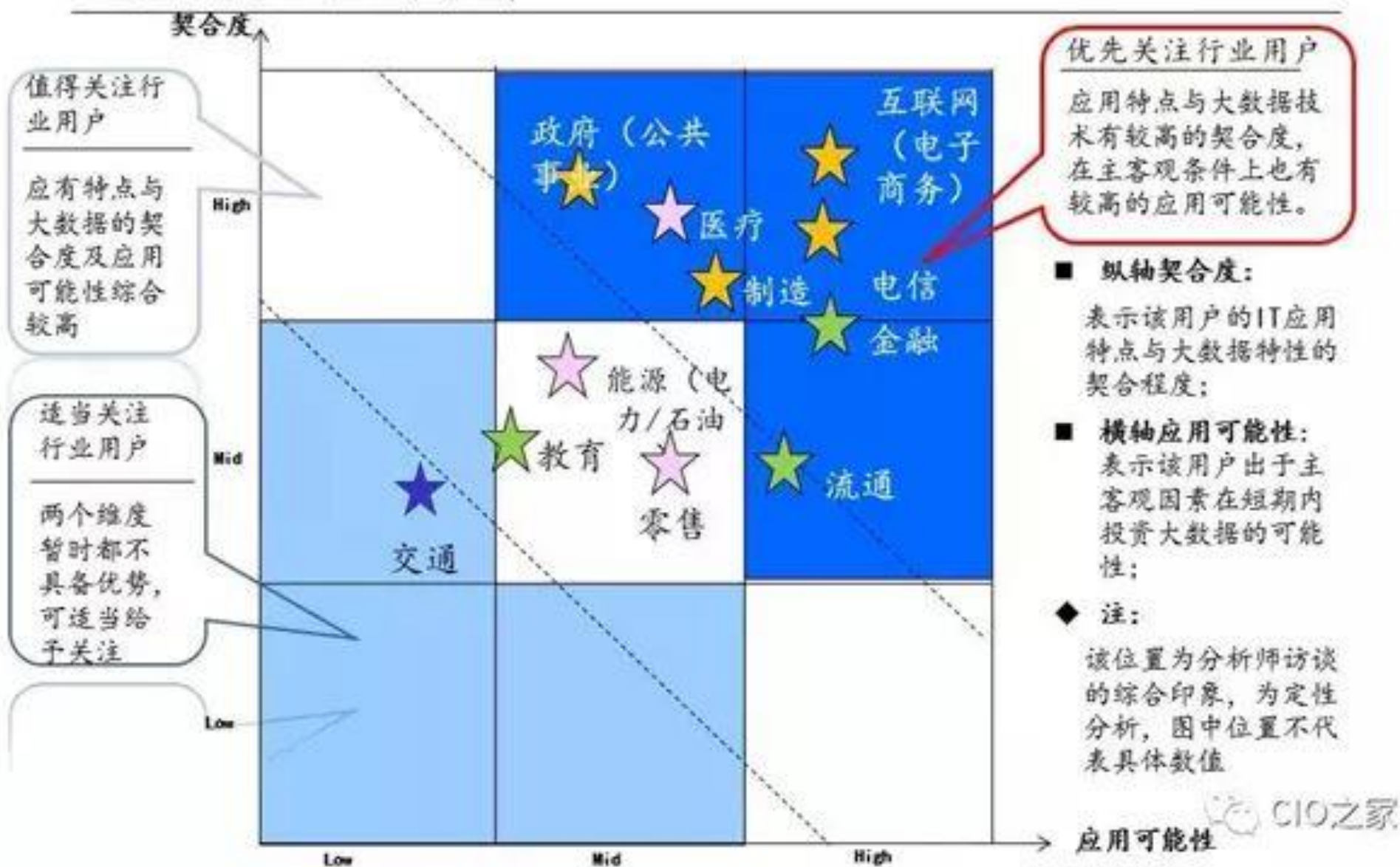
实施国家大数据战略

2017年12月8日，中共中央政治局进行第二次集体学习，听取北京理工大学副校长、中国科学院院士梅宏就这个问题的讲解，习近平总书记指出：

- 大数据是信息化发展的新阶段
- 推动大数据技术创新发展
- 构建以数据为关键要素的数字经济
- 运用大数据提升国家治理现代化水平
- 运用大数据促进保障和改善民生
- 切实保障国家数据安全
- 领导干部的基本功
 - 善于获取数据、分析数据、运用数据，是领导干部做好工作的基本功。各级领导干部要加强学习，懂得大数据，用好大数据，增强利用数据推进各项工作的本领，不断提高对大数据发展规律的把握能力，使大数据在各项工作中发挥更大作用

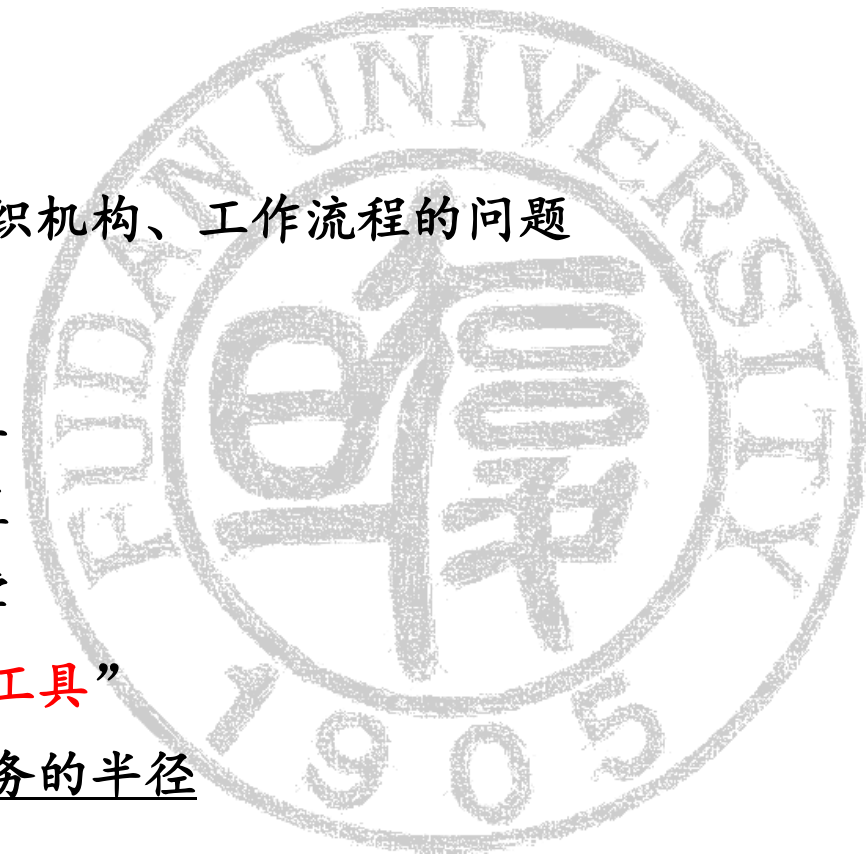


大数据行业应用分析



教育始终在苦苦追求着大数据

- 当前的现状 ——但获得回报依然艰难
 - 已建大量IT系统
 - 庞大的组织架构
 - 复杂的流程制度
 - 烟囱方式的孤岛
 - 多头的重复录入
 - 大数据系统平台过“大、重”，涉及太多改变现有组织机构、工作流程的问题
- 快速切入的需求(聚、通、用)
 - “集聚” --- 快速、有效、最大范围汇聚，实现全覆盖
 - “联通” --- 数据流通融合、共享交换，释放数据价值
 - “通用” --- 处理独立于应用系统，便捷提供数据支撑
 - 重视体验、突出“轻”字，从“**系统平台**”到“**简单工具**”
 - 关注关键的成功要素：刚需非增值、使用的频度、服务的半径



教育大数据的作用

为师生提供智能化服务

针对教师、学生、管理人员、行政人员全方面的业务和行为数据，利用数据采集、处理和挖掘技术，对教育大数据进行全面收集、合理清洗、精准分析、有效利用，实现用数据说话、用数据服务，以提升学校的智能化服务能力。



便于实现差异化教学，促进个性化学习

利用大数据，一方面，教师可以根据学生的不同情况推荐不同的学习资源，根据情况调整课程准备、教学顺序，真正做到因材施教；另一方面，大数据可以精细刻画学生特点，分析学生需求，为学生提供个性化的学习支持。

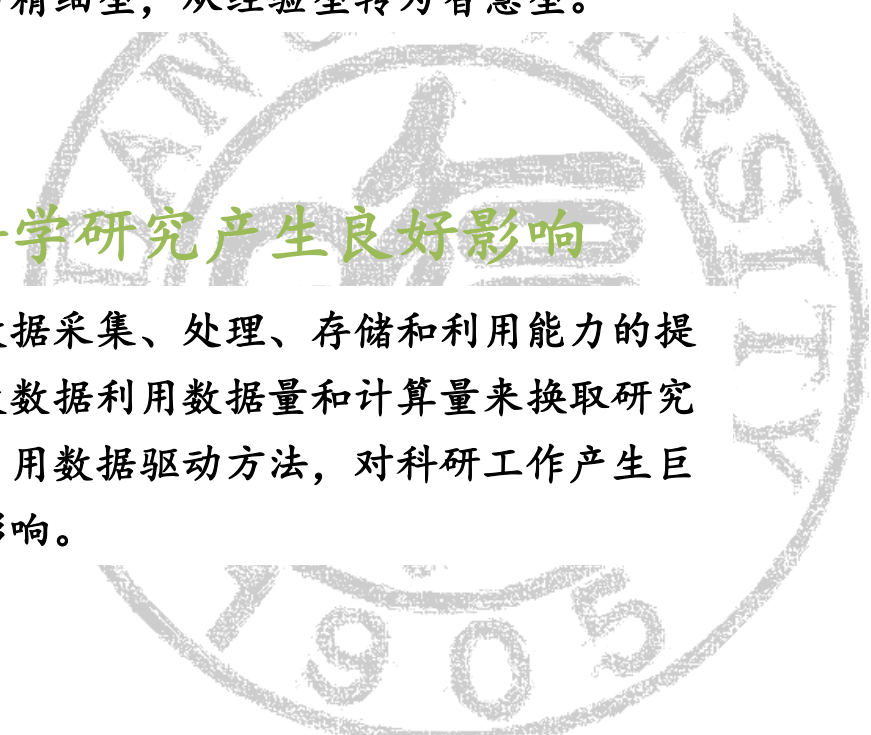


改进教育管理模式

利用大数据应用，高校管理者可以做到用数据管理、用数据决策，将过去的静态的、孤立的、滞后的数据变为动态的、交互的、实时的数据，将学校的教学科研管理，从粗放型转为精细型，从经验型转为智慧型。

对科学研究产生良好影响

随着数据采集、处理、存储和利用能力的提升，大数据利用数据量和计算量来换取研究时间，用数据驱动方法，对科研工作产生巨大的影响。



大数据的十大局限

➤ 一篇网上文章

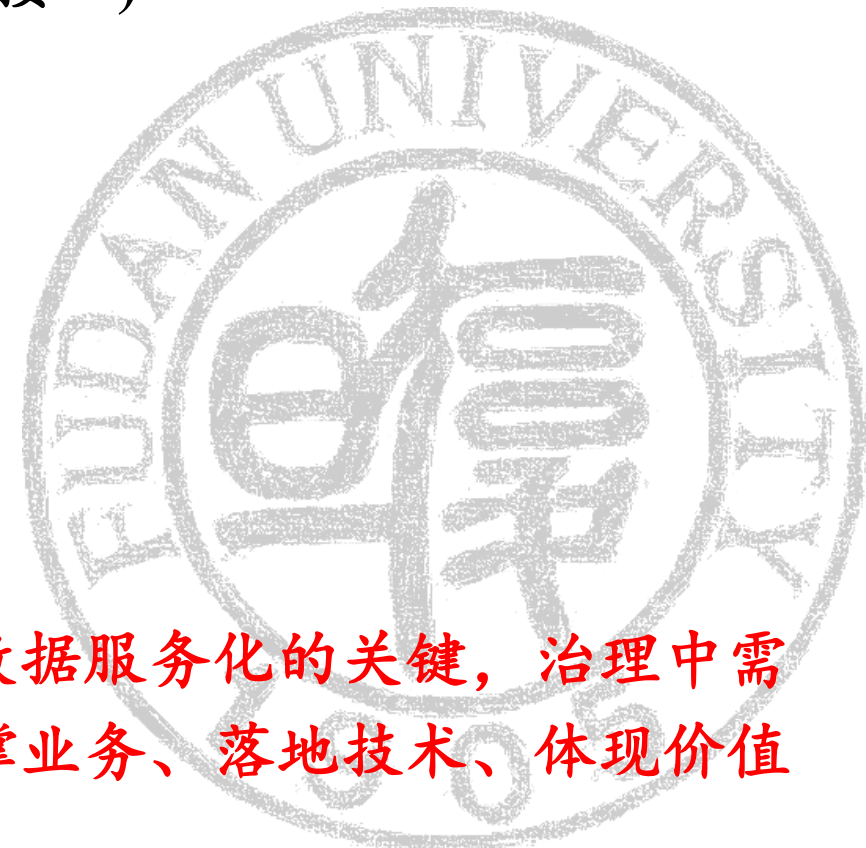
1. 数据噪声：与生俱来的不和谐(垃圾数据混杂)
2. 真实性：不得不接受的虚假(准确性无法保证)
3. 代表性：永不可能的全样本(物理和活动数据有盲区、代表性不足)
4. 完整性：广度与深度的缺失(孤岛导致两方面不足—网络和应用)
5. 时效性：秒级价值存在(信息更新加快，限制了应用的价值)
6. 解释性：不能没有因果关系(数据给出相关性，因果性需要思考)
7. 预测性：让过去决定未来(有很多未来靠过去经验预测有大风险)
8. 误导性：数据也会说谎(伪数据处处存在、包括状态监控)
9. 合法性：数据安全与隐私保护(必要的保护也会限制充分发挥)
10. 价值性：投入与收益的平衡(采/存/做/用的成本高昂)

强调：大数据代表只是信息而非智慧，思考才能发挥价值

军规 -- 大数据治理的12个技术原则

- 可视化管理数据资产(各种来源的数据、各种可视化手段)
- 自动获取数据信息
- 大数据服务智能化(热点数据预处理与缓存)
- 数据管理能力服务化(尽量能提供直接应用的API接口)
- 从需求开始控制数据质量
- 在集成点检查数据质量
- 持续积累检核规则
- 自动化质量评分
- 管理核心数据定义
- 为数据标准添加技术属性
- 识别与管理业务语义
- 关联业务语义与技术

大数据治理是大数据服务化的关键，治理中需
转变管理定位，支撑业务、落地技术、体现价值



撬动“大数据”：技术很重要，但思维是关键

- 从“流程”核心 ---》 “数据”核心
- 由功能是价值 ---》 数据是价值
- 从抽样 ---》 需要全部数据样本
- 由关注精确度 ---》 关注效率
- 由因果关系 ---》 关注相关性
- 从不能预测 ---》 可以预测
- 从人找信息 ---》 信息找人
- 由人懂机器 ---》 机器更懂人
- 由生产产品 ---》 用户定制

大数据思维是客观存在的、新的思维观：用大数据思维方式思考问题、解决问题是当下的潮流，开启了一次重大的时代转型



使用大数据提供优化的用户体验

大数据是帮助提升用户体验的主要工具，随着分析技术的发展和普及，可以利用各种来源的复杂数据集，帮助我们更加清晰地了解用户行为、进而进一步提升用户服务

- 提高对目标用户的理解
- 查看完整的用户历程
- 定制用户体验
- 提供多渠道支持
- 帮助用户节省时间



未来的智慧校园



大数据

数据的收集和处理

人工智能

数据的深度学习和利

云计算

数据的存储

物联网

数据的来源

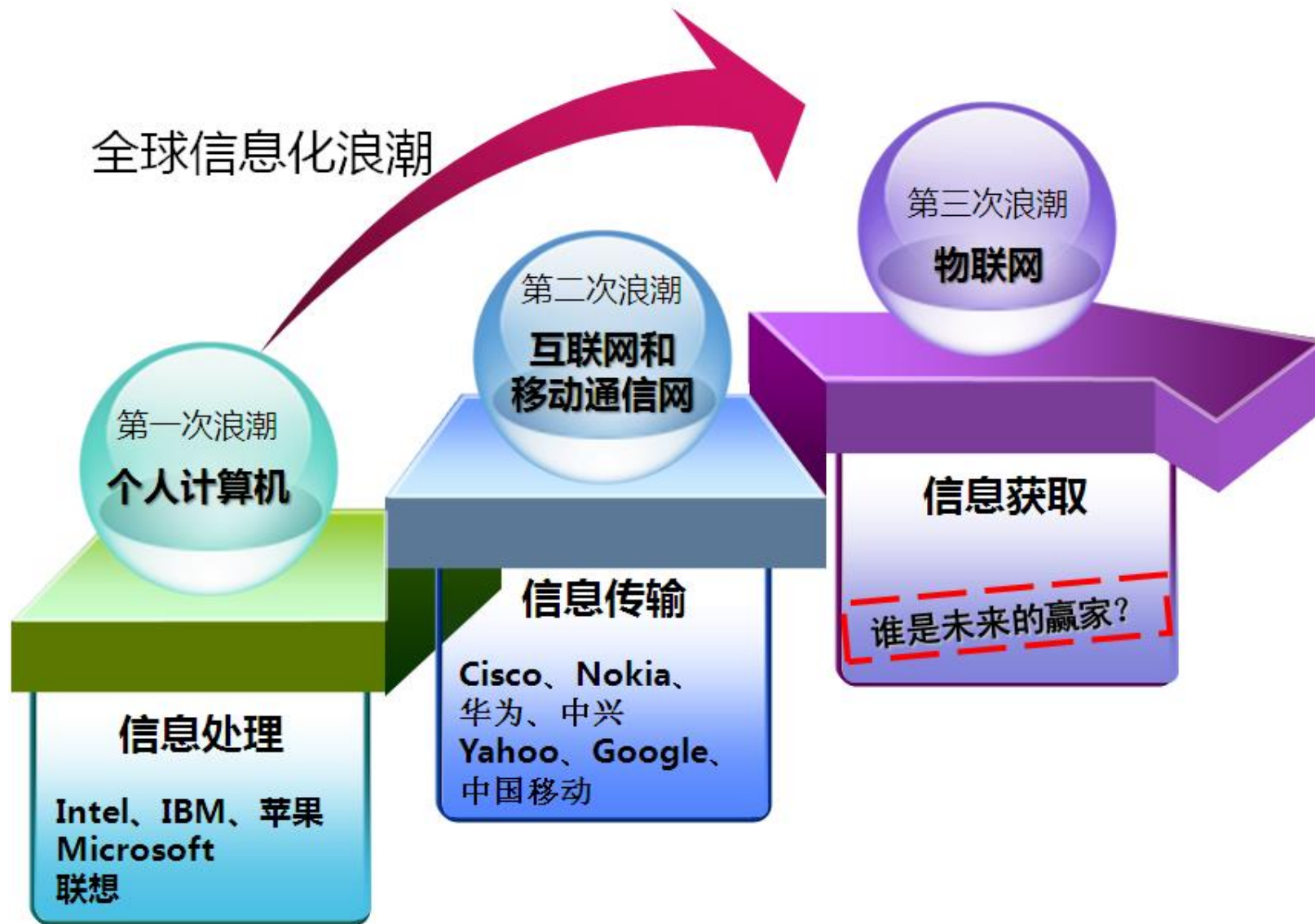
未来的智慧校园，不会由孤立的技术来实现，而是融合了各个学科门类的前沿技术，大数据的应用也只是其中之一

(智)慧：人的智能 + 人工智能 → 要培养提出问题的能力

人的创造力不同于机器的创造力、相互间是互补关系，要有人工智能思维，思考利用人工智能如何更好地帮你“做事”

综合能力的评价关心的是短板，而进入社会后的竞争更考验的是长板(特长)

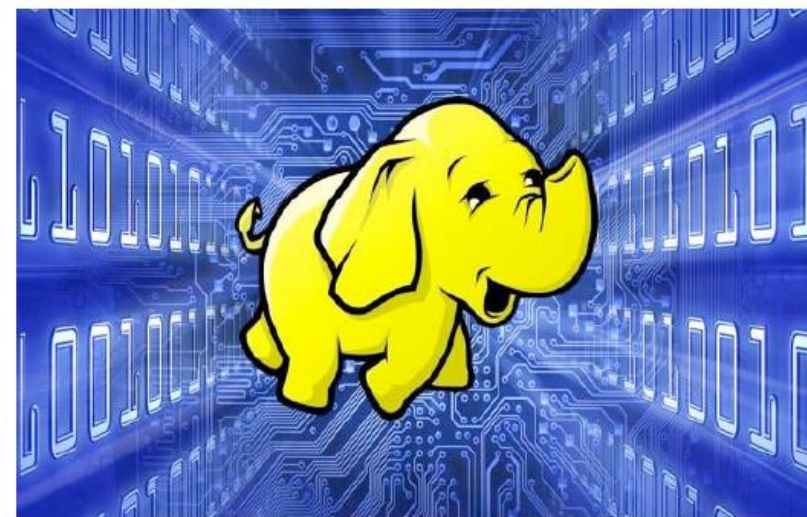
物联网时代将带来全球第三次信息化浪潮



智慧体现

- 更透彻地感应和度量世界的本质和变化
- 促进世界更全面地互联互通
- 事物、流程、模式、运行方式更深入的智能化

随着信息技术的不断发展和成熟，人们可以基于物联网来全面感知和获知万物信息，应用大规模计算机集群以及云计算技术去存储、加工、建模、预测和分析所得到的海量大数据，从而最大化地利用信息资源，并从中获得前所未有的洞察力，从而更快、更好地作出决策



全面的日志利用

从技术和业务的角度设计网络数据采集与分析工具

查询

- DHCP分配查询 (有线/无线)
- 认证信息查询 (有线/无线)
- DNS请求查询
- 代理服务器日志查询
- NAT地址转换查询
- VPN认证数据查询
- 防火墙日志查询

个人

- 行动轨迹 (个人)
- 网络行为轨迹
- 个人网络实时接入信息
- 校园网关系
- TopN排行

区域

- 楼宇 (区域) 当前用户数
- 楼宇 (区域) 网络质量
- 楼宇 (区域) 在线用户
- 楼宇 (区域) 网络流量
- 楼宇 (区域) 异常行为
- 楼宇 (区域) 平均网络延时
- 楼宇 (区域) 干扰统计

专题

- 图书馆过量下载
- 图书馆区域流量分析
- 校医院区域流量分析
- 学生活动场所流量分析
- 代理服务器使用统计
- 网络出口流量汇总
- 网络主干流量汇总
- 无线网掉线汇总
- 二级域名访问汇总
- 班级/课程学生上网行为汇总

数据源

- VPN采集 (配置) 监控
- DNS采集 (配置) 监控
- NAT采集 (配置) 监控
- DHCP采集 (配置) 监控
- 流量采集 (配置) 监控
- SYSLOG采集 (配置) 监控

报表

- 校园网人数统计报表
- 校园网出口带宽统计报表
- 校园网无线统计报表
- 校园网域名访问统计报表

采集

- VPN采集配置
- DNS采集配置
- NAT采集配置
- DHCP采集配置
- 流量采集配置
- SYSLOG采集配置

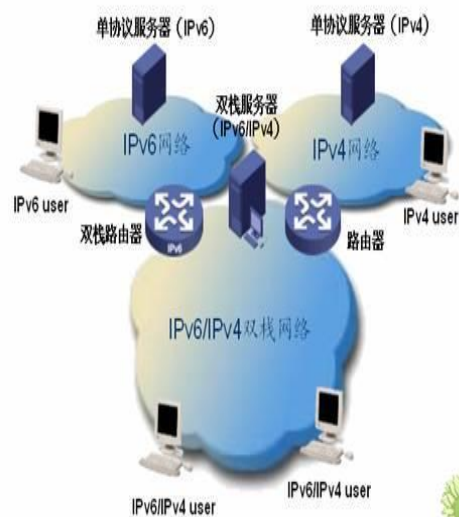
系统

- 用户管理
- 授权管理

未来的技术发展趋势

- 通信网络：IPV6、4G、5G、跨层结构
- 计算机与服务器：瘦客户端 + 高速宽带网络 + 云计算
- 物联网：RFID、无线自组织网络、传感器技术
- 大数据：存储、分析、建模及预测

物联网 + 云计算 + 大数据 = 智慧地球



思考

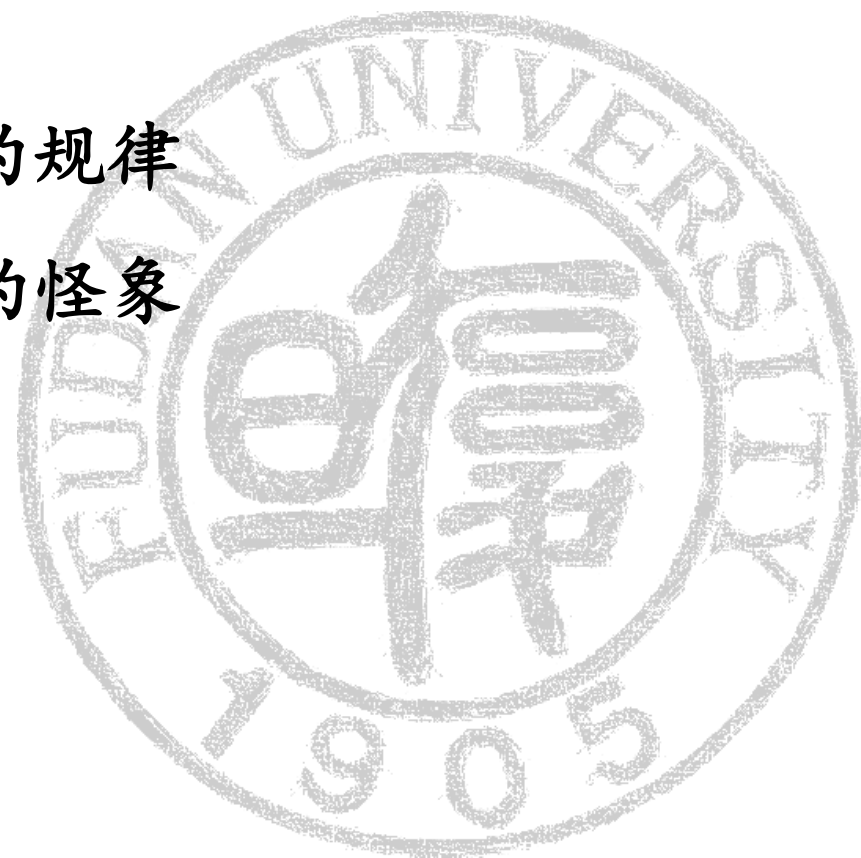
- 互联网技术迅猛发展，数据采集更加便捷、迅速、完整、真实
- 不但需要“大数据”、还需要“正确数据”
- 在实际应用中要避免伪数据、避免“简单粗暴”判别方式

基于大数据、互联网+ 的智慧校园前景远大!



思考

- 信息采集在于“广”
 - 跳出显像业务的范畴、要有数据“大视野”
- 数据挖掘在于“深”
 - 意义不在数据本身、而是隐藏在背后的规律
 - 要尽力避免“数据丰富、信息贫乏”的怪象
- 安全保障在于“严”
 - 个人隐私信息保护
 - 移动终端安全保障



勿忘初心 ~ 信息化的目标

- **不要为信息化而信息化**
 - 信息化应该促进IT技术支持的服务延伸和提高
 - 虽离不开网站和软硬件系统、但绝非仅此而已
- **信息化规划与实施方案**
 - 信息化规划需高屋建瓴、顶层设计
 - 具体实施规划则需能踏踏实实落地
- **数据治理与大数据应用**
 - 顶层设计是之上而下的
 - 实际应用是底层启动的

有温度的信息化才能使信息化支撑升级到信息化赋能!



Gartner: 2020年十大数据和分析趋势

几年前大数据的诱人前景，及最近机器学习和其他类型人工智能的潜力，推动了数据和分析技术在组织中的吸引力；这是因为数据和分析在数字业务中发挥着越来越大的作用，已经成为了实现关键功能的关键部分

据此，Gartner IT 提出了“将改变你的业务的10个数据和分析趋势”，这些趋势符合三大主题：智能、新数据格式、规模

增强分析

增强数据管理

NLP(自然语言处理)/会话分析

图形化

商业AI/ML将主导市场而非开源

数据结构

可解释的AI

区块链

持续智能

持久内存服务器



没有最好、只有更好
需默默耕耘、耐得住寂寞
学会从无趣中找到真正的乐趣

有忙碌而有意义的工作是一种幸运，能愉悦而有成效地完成是一种幸福

诗在远方、路在脚下：要努力使幸运变成幸福！

谢谢！

不曾奢望卓越，无妨追求完美！