

教育部全国职业院校教学
工作诊断与改进专家委员会

16版高职状态数据平台 采集与应用若干问题

樊至光



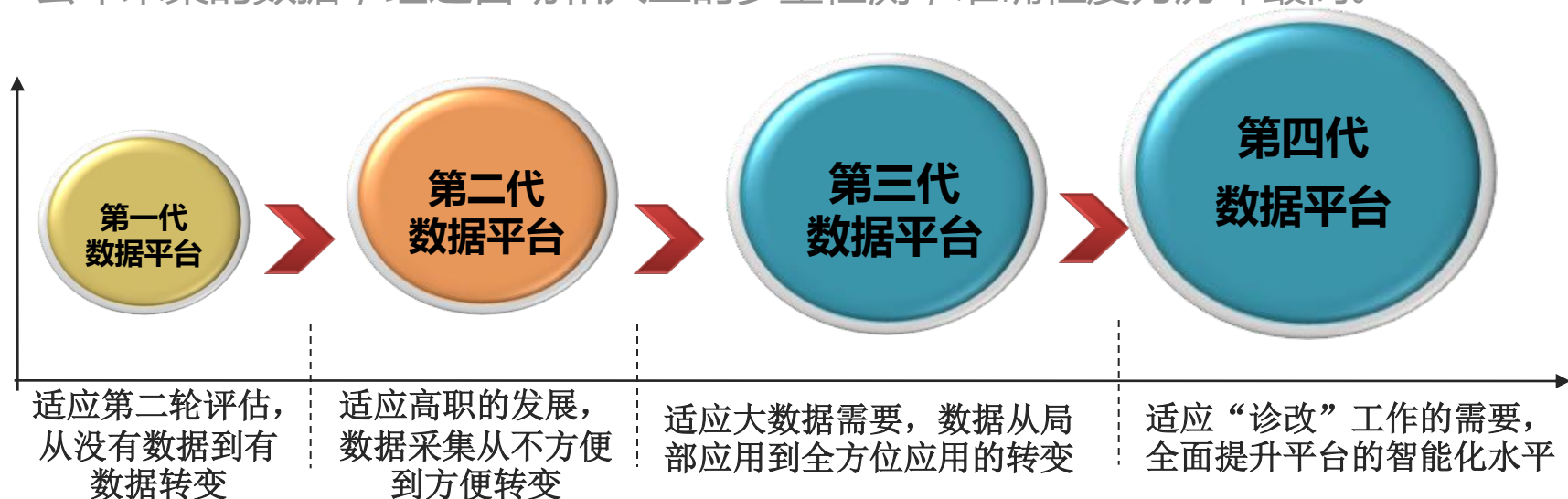
使命 采集 应用 展望



一、当前平台建设的两大使命

1、平台的发展历程

- “状态数据平台”建设从2008年启动至今，历时八年四个阶段，是目前高职院校众多数据平台中最能够直接反映学校“人才培养状态”的、最具应用价值的数据库体系，至今，数据平台覆盖几乎全部独立建制的高职院校，积累到的数据记录高达数十亿条；
- 从2014年起，“平台”升级为“数据采集与管理”系统，将数据的应用放到非常重要位置，目前可直接进行对比应用的数据已经有四年；
- 去年采集的数据，经过自动和人工的多重检测，准确程度为历年最高。





平台设计的三大思维空间：

**需要那些数据？
如何有效采集？
怎样充分利用？**

基本结构

数据体系

数据的规模
数据的定义（标准）
与其他数据体系的关系
数据之间的逻辑关系

数据采集系统

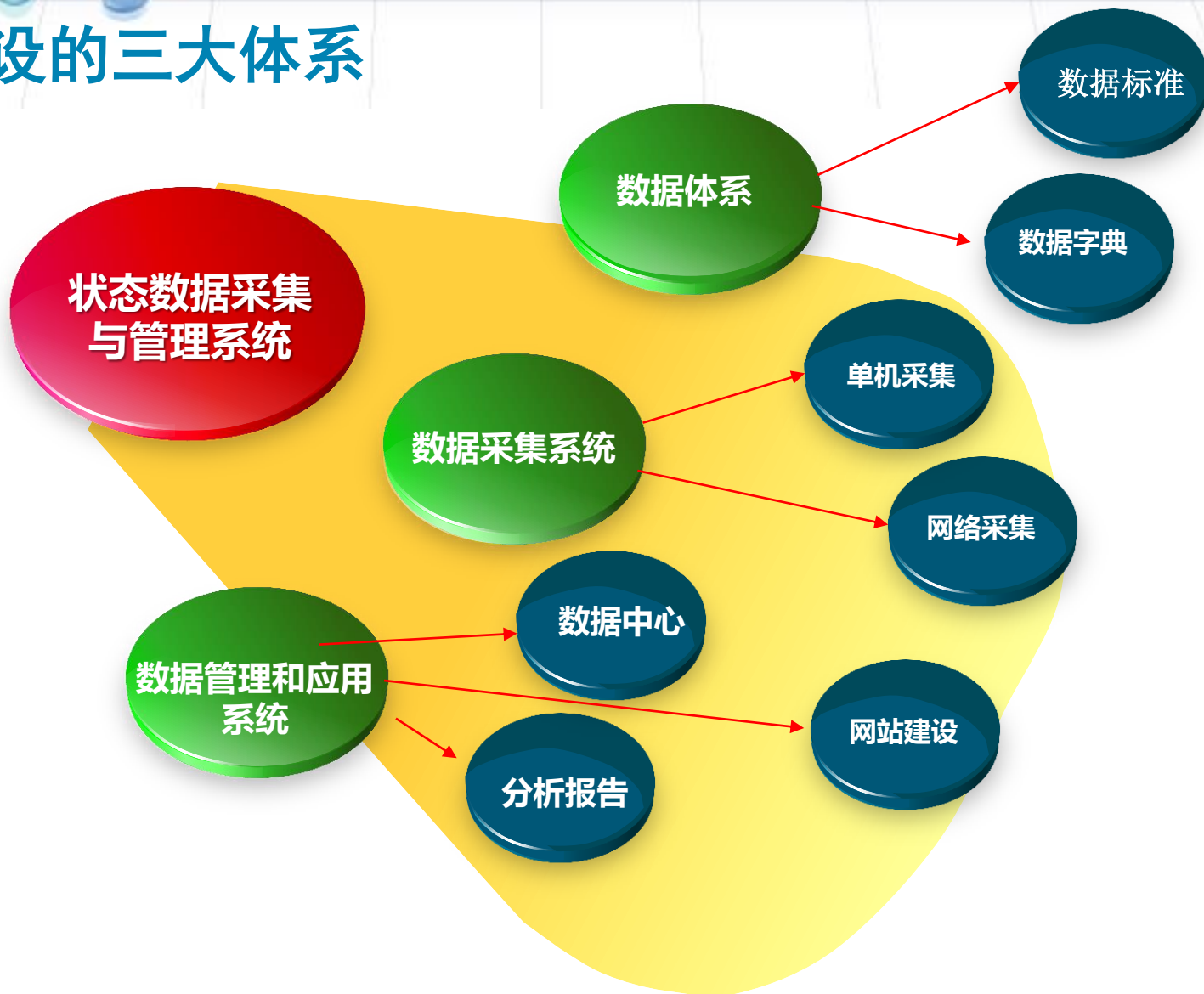
数据采集如何优化
如何让数据真实和准确
配合“诊改”工作，“过程采集”如何落实

数据应用系统

支撑“诊改”工作，
数据的深度挖掘
和多层面应用
数据的综合分析
移动应用的探索

状态数据采集 与管理系统

平台建设的三大体系



2、当前平台建设的两大使命

➔ **诊改工作的数据支持**

➔ **与国家金教工程对接**

使命一：诊改工作的数据支持

教育部有关诊改工作的“2号文”和“168号文”都明确指出：要形成基于职业院校人才培养工作状态数据的诊改工作机制，保证职业院校人才培养质量持续提高。说明状态数据平台建设，是诊改工作的基础性建设，是检验诊改工作是否有成效的关键之一。

文件还提出了各个学校应以“人才培养工作状态数据”应用为中心，推动学校的教育教学管理信息化全面上水平。

因此，状态数据的“深度应用”和“高效采集”将是学校建设的重要部分。

使命二：与国家“金教工程”对接

- 国家“金教工程”项目，是由教育部于2007年开始启动的，旨在建立和完善全国教育系统信息化公共服务和管理体系的重要工程；
- 工程重点之一，是要建立**教育基础信息数据库**，为教育管理与决策提供科学依据；
- 高职“状态数据”将作为“金教工程”数据体系中高职数据的主要组成部分，数据的完整性要求更高。

教育大数据应用的特点

- 因此，无论是“诊改”工作，还是“金教工程”，都从不同的侧面对“状态数据”提出了更高的要求，其重点都是要推进教育“大数据”的全面应用。

- **数据覆盖面更广量更大**
- **数据的实时性要求更高**
- **数据的真实性要求更强**

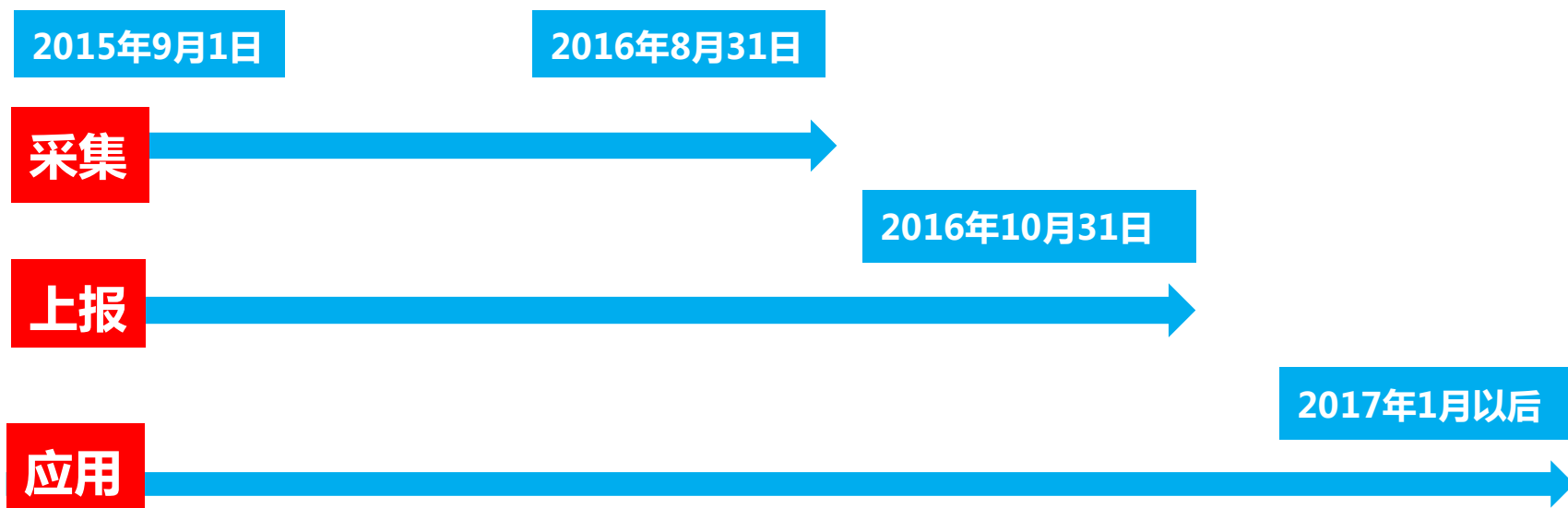


3、当前数据采集主要问题和解决原则

- **突出问题：**
- 数据**时效性**较差，数据应用受到制约；
- 数据填写量较大，不仅数据的真实性无法保证，数据采集难度也不少；
- 数据的覆盖面还有限，还不能反映学校的全面状况；
- **解决原则：**
- 强化数据的“源头”、“过程”和“自动”的采集；
- 解决“**实时**”、“**真实**”和“**方便**”等问题。



学生数据滞后情况



作为平台最基础的学生数据，从数据采集到数据应用，时间滞后了两个年度。



二、数据如何采集？

1、今年的平台将“过程采集”作为设计重点

- 今年的采集平台，既要与往年的“年度采集”衔接，完成2016年度的数据采集任务（所有学校），又要配合全面铺开的“诊改”工作，启动数据的“过程采集”（实时采集），实现平台数据更加“实时”的目标（部分学校）。所以，今年的数据采集，是承上启下的关键一年。
- 今年的采集系统，一共推出**五个不同版本**，是历年来版本最多的，请大家一定要注意分辨。
- 根据全国专委会的要求，与“诊改”工作相配合的“过程采集”或“实时采集”，将是今年平台设计的重点和应用趋势，尽管今年不是所有学校都马上应用。

今年平台五大版本

1. 原版本优化用于，年度采集

标准版

3. 全新版本，实现过程（月度）采集

过程采集A版

网络版

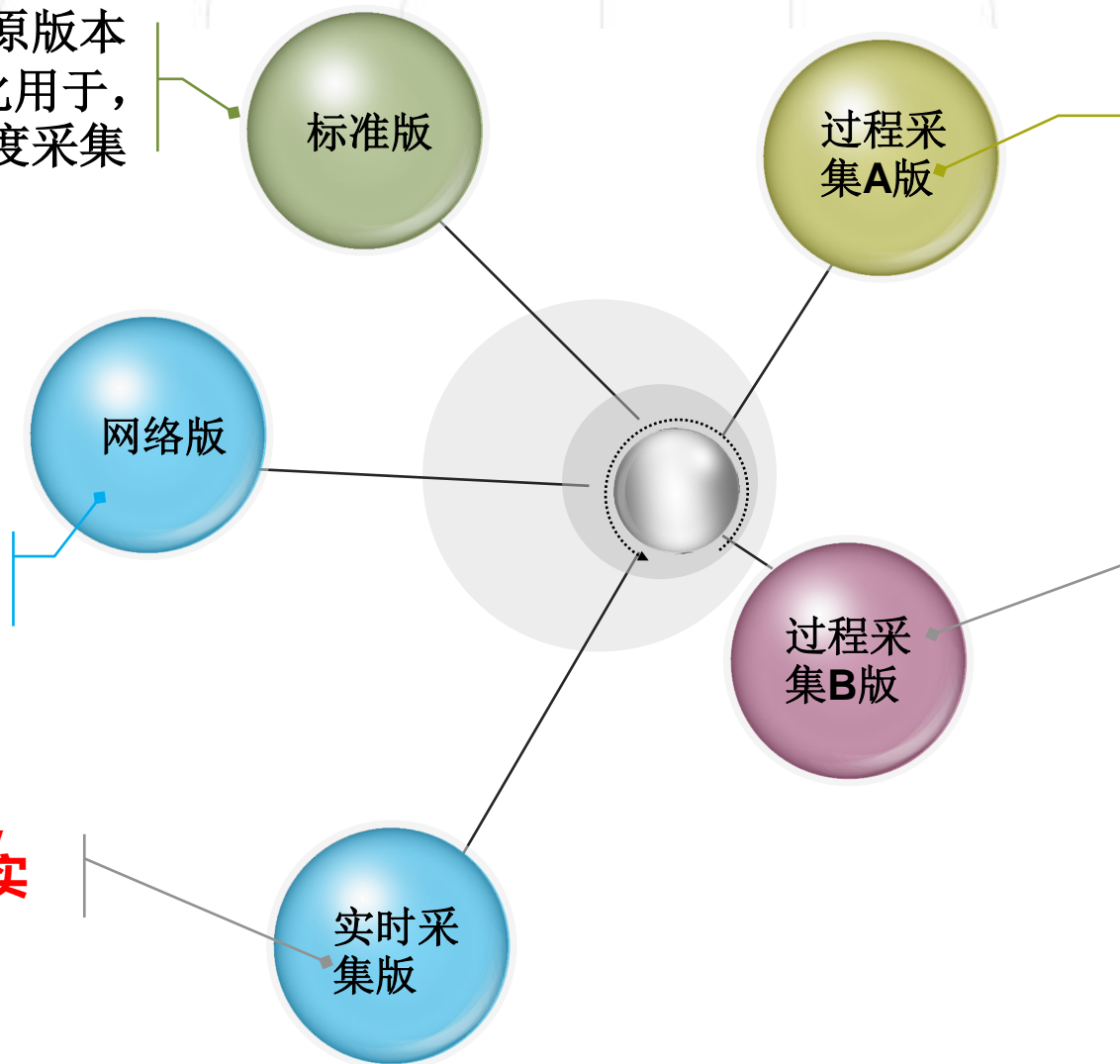
2. 原版本优化，用于年度采集

4. 全新版本，实现过程（月度）采集

过程采集B版

5. 全新版本，实现过程（实时）采集

实时采集版



五种采集方式比较

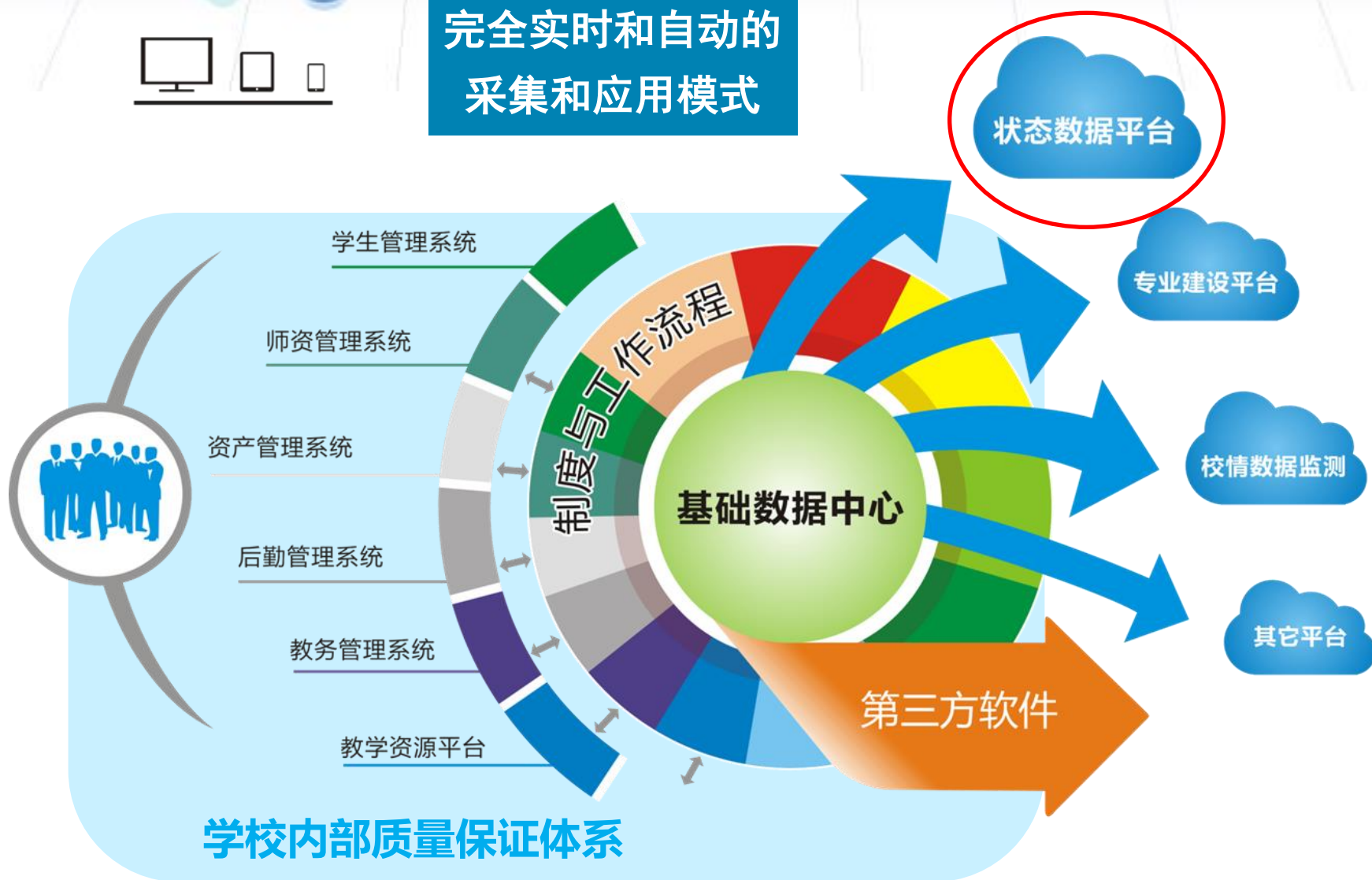


采集方式	采集方式	过程数据	与学校系统的关系	还需年度采集?	诊改作用	选择建议
年度采集	手工完成	无	无关	需要	一般	过渡方案
月度采集	手工或自动完成	有部分	无关或有关	不需要	较好	过渡方案
实时采集	自动完成	有完整	关系密切	不需要	很强	推荐方案

推荐方案：“状态数据”与“学校内部质量保证体系”自动对接



完全实时和自动的
采集和应用模式



2、今年（年度）数据采集的主要特点（重点）

- 1、扩充平台的数据量，将在校学生数据全部纳入平台管理
- 2、字段重新定义和分类管理，数据采集难度下降
- 3、完全实现数据源头唯一性，计算结果不能修改
- 4、发展性数据也从历史数据中比较计算而来
- 5、采集数据全员参与，各人对自己数据的真实性负责
- 6、进一步强化数据内部之间的关联性
- 7、加强了系统的检错与纠错功能，明显有问题的数据不能上传
- 8、对同一指标的不同理解，自动计算不同结果并如实展示
- 9、无论以何种方式采集数据，最终都经“网络版”统一上传
- 10、重新修订《数据字典》以统一数据标准
- 11、自动检测并提示更新版本，不更新将无法使用



2016版” 状态数据采集与管理平台” 调整优化情况汇总 (不含案例分析部分)

	指标	字段	说明	注释
增加	4	28	14	11
删除	1	1	0	1
修改	9	8	2	1

仅“字段”的变化

	字段	数量
增加	所在城市、培训对象（特定群体、小微企业）、留学生数（境外、国外）、(预科生、进修生、成人脱产班学生数合计)、夜大（业余）学生数、函授生数、绿化用地面积、一卡通使用、无线覆盖情况、财政拨款总额、政府购买服务到款额、技术服务到款额、是否行政班授课、授课学期、课题分类、著作与论文分类、是否国际合作专业、生源地、来自军队、招生方式、是否常住户口在农村、是否订单（定向）培养、是否建档立卡贫困家庭、年级、学制、入学日期、是否毕业、是否对口	28
删除	一体化教室	1
修改	培训规模，校舍总建筑面积，接入互联网出口总带宽，鉴定数（人次天），有无高校教师资格证书，生源类型，订单（定向）培养情况，捐赠项目名称	8

今年平台数据扩容主要是学生数据

- 今年平台数据容量的扩充，主要是将所有在校学生数据全部纳入平台，进行统一的数据管理与应用。
- 第一、在校学生数据是学校数据体系的基础，是学校质量管理的依据；
- 第二、在校学生数据是状态数据平台大部分数据自动生成的重要源头；
- 第三、在校学生数据一般在学校录取新生时已经生成，学校只需通过模板批量导入，难度并不大。



- 将平台所有数据重新定义和分类管理

字段总数	2217	100%
系统计算	1186	53.5%
自动获取	423	19.1%
源头采集	608	27.4%

- 将“源头数据”细分为“动态”和“静态”两类

源头采集		608		27.4%			
其中	人工采集	可从往年数据中提取		275	45%	12.4%	
		合计		333	55%	15%	
		其中	年度变化	282	84.6%		12.8%
			学期变化	11	3.3%		0.5%
			月度变化	40	12.1%		1.8%

绝对静态

相对静态

- “动态数据”才是最活跃和最有意义的部分

源头采集		608			27.4%			
其中	人工采集	可从往年数据中提取			275	45%	12.4%	
		合计			333	55%	15%	
		其中	年度变化	282	84.6%			12.8%
			学期变化	11	3.3%			0.5%
			月度变化	40	12.1%			1.8%

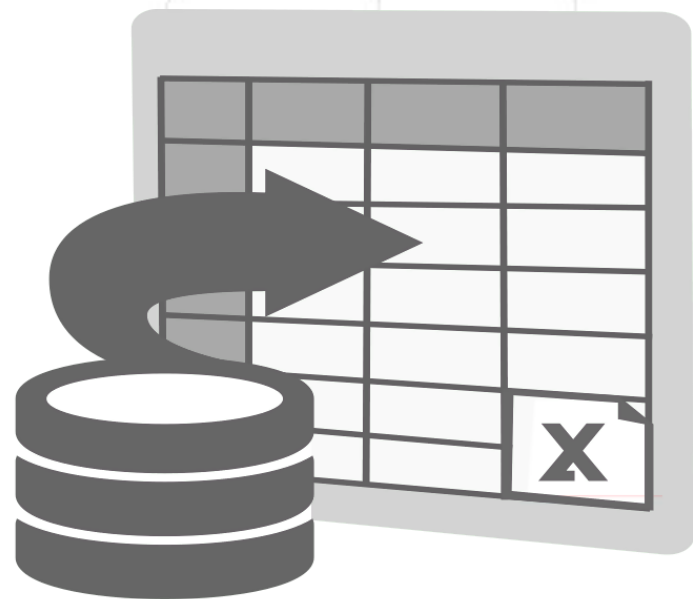
动态数据





- 将数量有限的“**动态数据**”作为“关键”数据，平台数据的“过程采集”和“实时采集”将成为可能

字段总数	2217			100%			
系统计算	1186			53.5%			
自动获取	423			19.1%			
源头采集	608			27.4%			
其中	人工采集	可从往年数据中提取			275	45%	12.4%
		合计			333	55%	15%
		其中	年度变化	282	84.6%		12.8%
			学期变化	11	3.3%		0.5%
			月度变化	40	12.1%		1.8%



教师授课数据可以批量导入



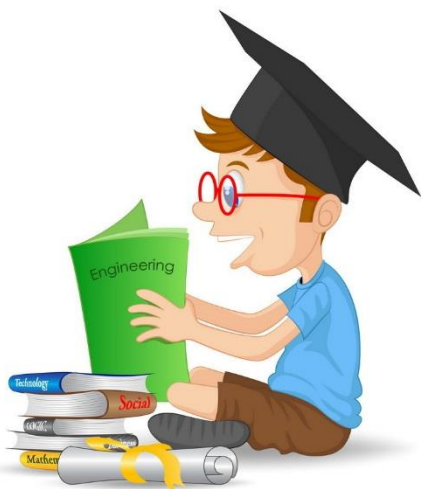
但须交由教师本人自己确认



数据主要源头



授课情况
学历情况
职称情况
科研情况
等等



评教情况
顶岗实习
就业信息
等等

今年实施

下一步实施



3、工作进程

学校情况	采集方式	启动时间
大多数学校	年度采集	延续以往的做法
试点学校	月度或实时采集	今年11月
信息化条件较好学校	实时采集	自定

计划用三到五年时间，全部学校进入“过程采集”

预计时间表

年度	时间	工作内容	工作主体
2016	7月	发布2016版平台	专委会
2016	9月	发布“过程采集”版平台	专委会
2016	10月底	上报2016年度状态数据	所有学校
2016	11月	“过程采集”版平台使用培训	专委会
2016	12月	开始进入月度（或实时）采集	试点学校
2017	3月	试点院校实验性过程采集工作总结交流	专委会
2017	11月	将“过程采集”学校的数据汇总上报	试点学校
2017	12月	新一轮“过程采集”开始	参与学校

通过“过程采集”或“实时采集”的数据才具有以下特点：

- 源头唯一
- 过程采集
- 全员参与
- 数据共享



- 这样的数据才符合“诊改”工作的基本要求
- 学校可以把握信息化建设的契机，结合数据采集方式的转变，从根本上解决数据滞后问题，让数据变得鲜活和真实，学校将进入全面的数据管理的新时代。

4、上传后的数据检测与完善

- 各学校上传数据后，务必关注11月份“数据平台”给每个学校反馈的数据检测报告（可登录平台查看），发现问题及时修改并重新上传；
- 12月关闭平台后，学校需要修改数据，需要经过教育主管部门批准，并记录在案。

相关建议

- 今年学校的“状态数据”采集，希望尽可能使用“网络版”；
- 在推进学校信息化建设的基础上，尽快建立学校内部质量保证体系数据系统，为与“过程采集”数据整合做准备；
- 学校在部署学校的业务（管理）系统时，应同时考虑好与采集平台的对接，或者实现在学校已有业务（管理）系统中自动抽取数据，使学校的信息化与数据应用同步；
- 试点学校外的其他学校，尤其是条件较好的“示范”和“骨干”院校，可根据学校的信息化发展状况，率先进入“实时采集”环节，专委会将提供理念和技术方面的支持。



三、数据如何应用？

1、已建成多层次应用的高职院校数据中心

中国高职院校数据监测中心

新功能 问卷调查 进入页面

全国高职院校分布图



全国高职院校在校生分布图

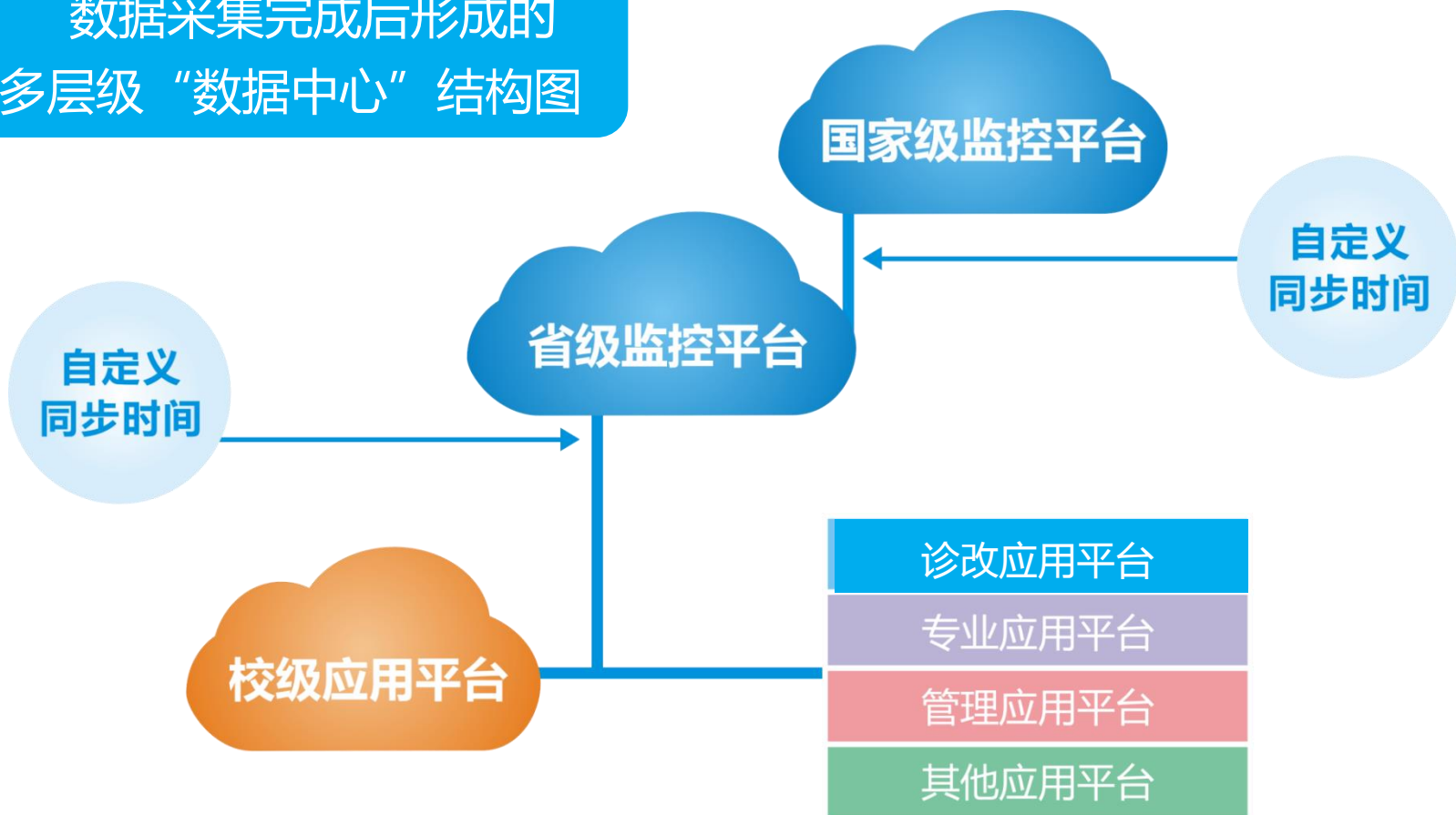


全国高职院校校布点图

- 65所以上
- 49-64所
- 33-48所
- 17-32所
- 0-16所



数据采集完成后形成的
多层次“数据中心”结构图

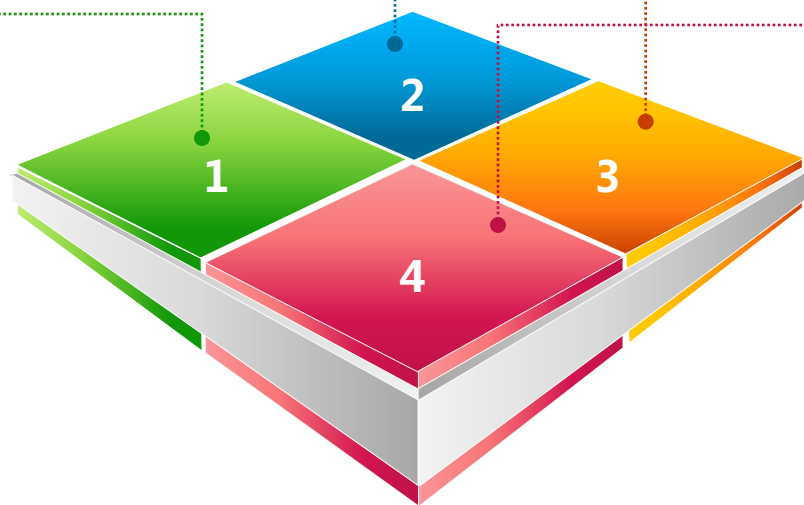


服务于
各学校

服务于
教育专家

服务于
省教育厅

服务于
教育部



高职状态数据中心

2、数据中心的特点

- 汇总了2012年、2013年、2014年、2015年的全国数据，平台已经有4年的数据积累；
- 所有数据相互关联直至数据的最底层，即从全国数据到省级数据再到学校数据的完全贯通，教育部一键直达每一所学校（省级平台也是如此）；
- 几乎每一条数据都实现了纵向和横向的全方位比较，数据应用完整全面、脉络清晰；
- 在大多数的数据比较中，示范和骨干院校均为数据参照项；
- 大部分数据来自源头并自动计算，系统会对异常数据提出警示，但不会对数据做任何改动，以求客观真实；
- 系统会将自动过滤到的可能有问题数据，以邮件或网上问题提示方式反馈给学校，方便学校对自己的数据进行维护；
- 系统对数据的不断挖掘和组合将形成无尽的信息源。

3、校级数据中心在“诊改”中的作用

- 通过数据分析全面掌握学校人才培养质量状况；
 - 通过数据的全方位比较明确学校的发展定位；
 - 查询和比对对学校建设与发展有重要意义的的数据；
 - 强化用数据引导学校管理的现代化管理意识；
 - 提出新的数据需求扩充数据中心的应用面；
 - 在数据的应用中发现学校数据可能存在的问题。
- **可见，目前学校层面“数据中心”，主要是为学校教育教学管理和持续改进服务的**

学校发展坐标图

- 当前学校基本办学条件达标情况

- 学校所有数据历史比较（4年）

- 学校数据与示范、骨干、同类、全省、全国学校的全方位比较
- 学校在全国各中专项数据中定位

- 学校基础数据单项全国定位（排序）
- 全国情况相近（相邻）学校的参考比较

- 学校“发展指数”曲线比较
- 学校“发展状况”全国比较



平台深挖后的大量数据，对学校“诊改”很有意义

- 当前全国学生最多的专业是哪个？
- 全国最多学校办的专业是哪些？
- 全国就业最好或就业最不好的专业又是哪个？
- 那些专业能办，哪些专业不能办？
- 办某个专业需要准备哪些课程？
- 我校所办专业在全国同类专业中处在一个怎样的水平？
- 我校毕业生主要流向在哪里？
- 我校毕业学生起薪点在全国同类专业中处在什么水平？
- 我校与全国哪些学校比较相近？
- 我校与同类学校比较有哪些优势和不足？
- 等等

高职“状态数据”平台大数据显示（2015年度）

- **全国专业与学生情况。**
- 全国共有1230个专业，2015年招生的专业有1111个，平均每所学校29.47个专业，平均每所学校全日制在校生有6432人，平均每个专业218.26名学生（不含方向），平均每级学生只有70名；
- **全国专业规模情况。**
- 全国财务会计类专业学生规模最大，有在校生77.33万人，占全国在校生成数（827.21万）9.3%以上，全国有965所学校在办这一专业；其次是计算机类（学校开设最多的专业），在校生有51.43万，占6%；护理类学生数第三，在校生45.84万，占5.5%；三个专业的学生占全国学生总数的20.8%；



- **全国专业与学校情况。**

- 最多学校开设的专业分别为：计算机类，全国有1017所学校在开设，财务会计类排名第二，有965所学校在开设；第三名为市场营销，有915所学校开设，学生29.38万人；第四是工商管理，有842所学校开设，学生29.9万人；需要注意的，后三位的全部是财经大类的专业，也就是说，高职在校生中，财经大类专业的在校学生总数占20%。值得注意的是，这些专业毕业生的情况并不是最理想的，如在专业大类总数仅为79类中，就业率分别为15、36、63、43；就业起薪线也分别为38、61、51、41位。



- **学生就业最好专业情况。**

- 飞行技术、舞蹈表演、高速铁路信息控制、应用外国语四个专业应届毕业生就业起薪线分别为10000、5696、5000、5000，名列前四位。

- **同一专业毕业生在不同地区就业的差别。**

- 同样是财务会计类，在北京、浙江、江苏、广东就业，起薪线分别为3192、2531、2434、2404元（平均是2075元），但是，在广西、辽宁、青海、安徽，则分别只有1692、1705、1800、2076；首尾差距达到89%。

全国高职毕业生 就业平均起薪线 地区排名



	省份	起薪线
1	北京市	3142.45
2	西藏自治区	3000.00
3	河南省	2953.67
4	新疆维吾尔自治区	2790.00
5	浙江省	2770.07
6	广东省	2672.69
7	上海市	2640.73
8	湖南省	2631.11
9	江苏省	2615.13
10	江西省	2429.70
11	重庆市	2333.46
12	湖北省	2326.71
13	内蒙古自治区	2307.96
14	四川省	2295.64
15	陕西省	2244.88
16	天津市	2209.30
17	安徽省	2202.95
18	河北省	2183.53
19	海南省	2182.94
20	黑龙江省	2177.63
21	甘肃省	2175.63
22	云南省	2148.00
23	山东省	2139.47
24	福建省	2126.85
25	宁夏回族自治区	2060.40
26	贵州省	2035.70
27	吉林省	2017.46
28	山西省	1991.30
29	广西壮族自治区	1968.38
30	辽宁省	1939.69
31	青海省	1660.81
	平均	2334.65



- **专业与课程的情况**

- 全国专业平均开设课程30.68门，总体感觉比较均衡，但是通过数据深入挖掘发现，开40门课以上的专业占了28.4%，最高的前5位，高达350、318、195、179、165门；而仅仅开设10门课以下的专业，占11.6%，其中只开有1门课程的专业，竟然有911个之多。



- **人才培养计划的落实情况**

- 计划开设的专业课与实际授课情况相比，全国平均只完成了86.51%；全部完成计划的学校只有145所，仅占全部学校总数的11.25%；只完成计划一半以下的，也有56所学校。

- **通过比较这几年的数据发现，上述情况并没有多少变化。**

“状态数据” 的诊断效果与持续改进

- 一个学校，如果通过“状态数据”系统的多维扫描和诊断，学校的问题大致可以有一个基本轮廓；
- 当然，数据的逻辑判断受到很多因素的影响，如学校数据是否准确就是一个关键，所以检测结果只能作为参考，学校在自我诊断中对已有数据的分析和判断不可缺少；
- 与此同时，系统本身也在持续改进当中，各学校在“诊改”实践中需要怎样的数据体系以及如何比较，希望能够及时反馈给我们。



5、“数据应用”的重视程度还须不断加强

高职数据中心登录情况汇总

2015年7月到2016年5月

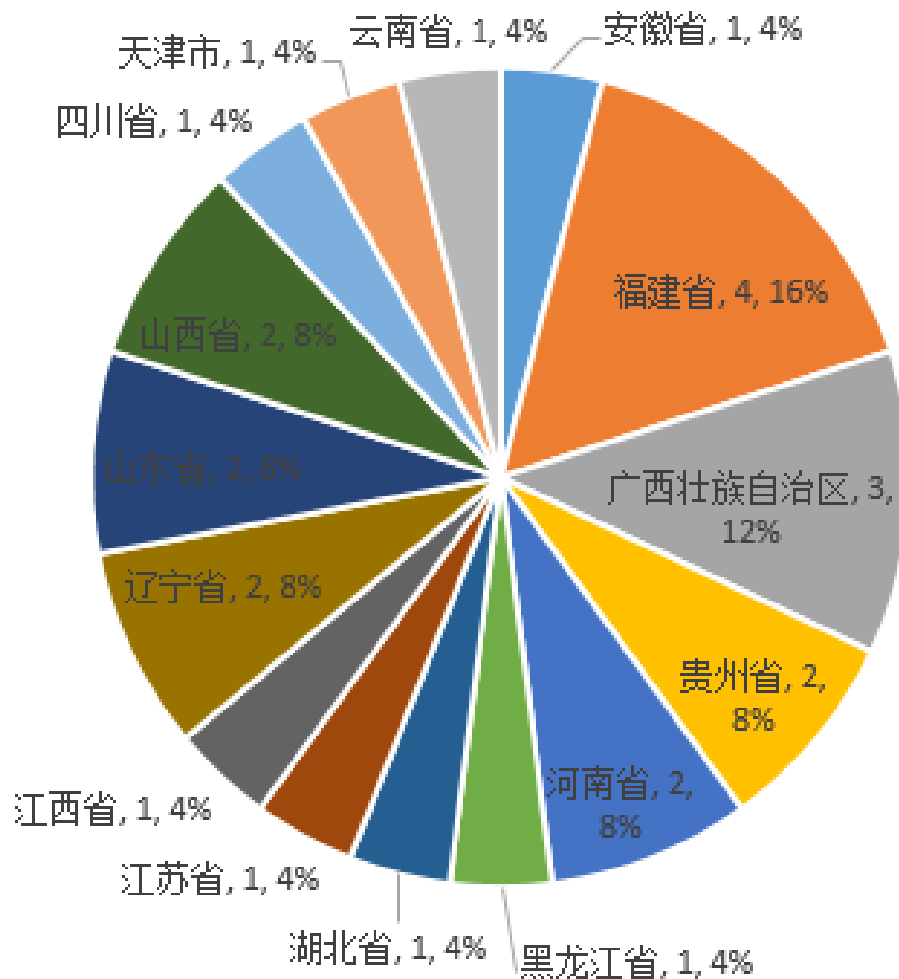
类别	总登录次数	平均登录次数	最高登录次数
总数	63625	45	588
学校	58877	45	588
教育厅（委）	3527	110	446
教育部	295	59	246
专家	826	12	226

从登录情况可以看到，学校平均每月登录4到5次，最高是每天一次以上，可见大部分院校登录“数据中心”的积极性较高。

问题一：登录情况很不均衡

- 平台数据显示：
- 有25所学校完全没有登录的；
- 仅登录1到3次的有37所学校；
- 合计62所学校。

未登陆分布



从“登录曲线”中也发现问题

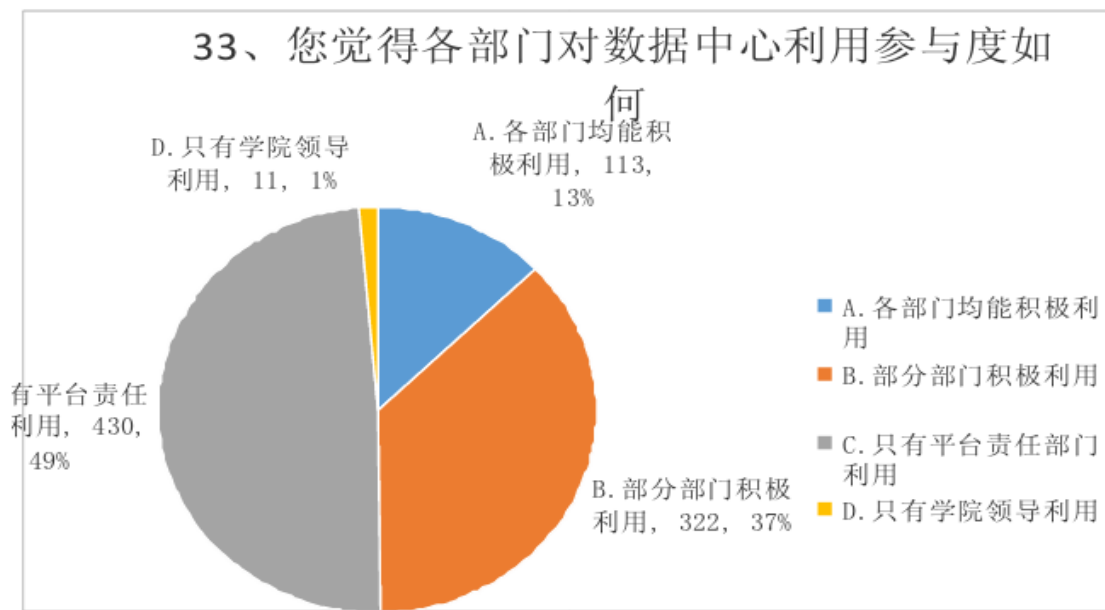


上述“登录曲线”说明，普遍学校登录的目的还是在于数据的填报，对学校乃至全国数据的分析和应用，还没有给予足够重视。

问题二：普遍学校对数据平台应用不到位

• 调查显示，谁在用数据中心的数据？

- A.各部门均能积极利用 B.部分部门积极利用 C.只有平台责任部门利用
D.只有学院领导利用



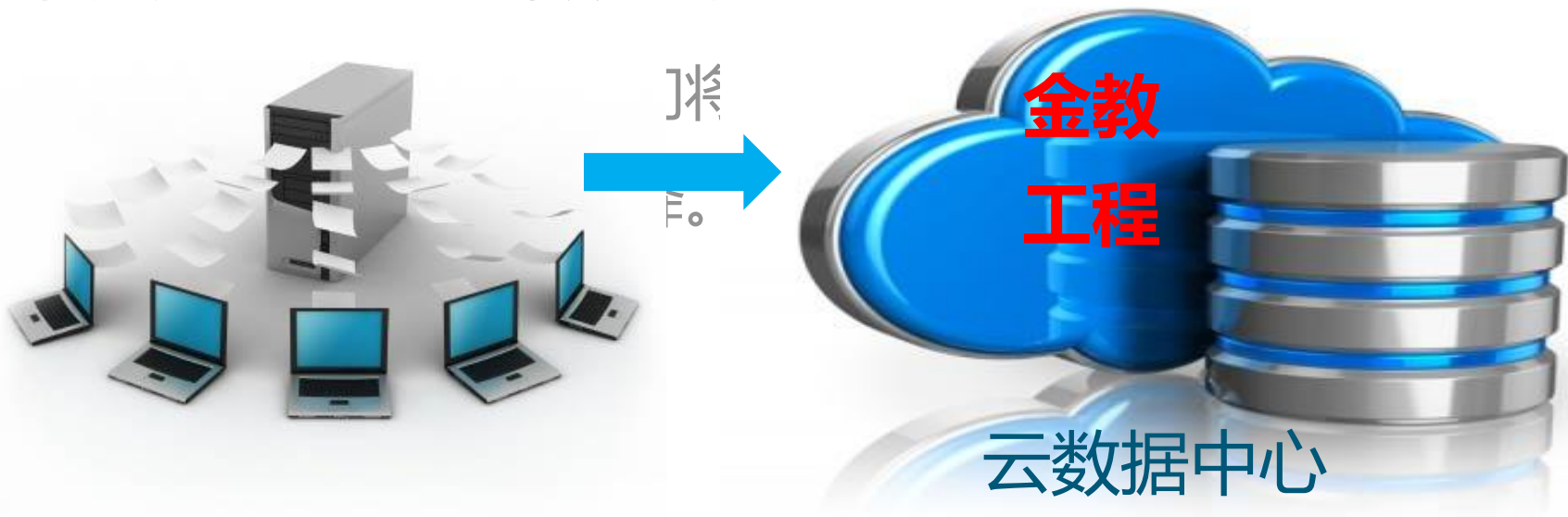
根据用户反映：除了责任部门，学校其他部门对“数据中心”的参与和应用程度普遍不高，尤其学校领导，关心不多、重视不够，对数据的真实性保障和深度开发应用都是很不利的。



四、平台建设前景展望

1、系统数据量必将大幅度增加

- 无论是“金教工程”还是“诊改”工作，对数据的需求将不断扩大，数据总量的增加是必然趋势；
- 但是，经过优化后的数据体系，将大幅减轻采集的压力；
- 与此同时，因为数据的主要应用者也包括学校，数据的采集将是为自己而工作，数据采集的积极性和主动性也将大大提升。



2、数据采集的“实时”和“自动”将是趋势

- 实时的数据才是真正意义的大数据（月度采集只是过渡）；
- 在工作过程中自动采集的实时数据，才是最真实、最可靠和成本最低的；
- 可以预料，“诊改”工作的推进，将与数据的实时化同步，也将与学校的信息化发展同步。



3、数据的多平台共享必将加强

- 教育大数据催生数据标准，数据的共享时代必将到来
- 当前，主要实现与《高职年度质量报告》数据共享
- 计划与“督导办”的能力评估实现数据共享
- 最终与学校各类数据平台（报表）数据实现共享



4、无论采集数据还是应用数据， 移动应用将是主流

- 需要了解什么信息，打开手机版数据中心问一下就知道。



5、 “数据中心” 的服务对象将进一步扩大



1 别的学校专业和课程如何？



3 企业需要的专业人才在哪里？



2 报考那个学校和专业呢？

6、真正的教育大数据应用亦将应运而生

考生检索那个学校比较多？
考生点击哪个专业比较多？
哪些地区的学生关注什么？
在校生的考生在信息关注上
有哪些不同？
老师较多关注哪些信息？
那些学校的学生和老师登录
较多？



非结构性大数据的收集、分析和应用意义深远

结束语：

- 状态数据管理系统将不仅仅是一年一度的数据采集工作，相比之下“日常应用”将变得更有意义；
- 各层“数据中心”以学校数据为基础、以全国大数据为背景全面展开，数据的意义将更为广泛和深刻，这些数据必将成为今后政府和学校科学管理和决策的重要依据，所有数据采集必须谨慎；
- 希望学校以“过程采集”和“数据应用”为契机，全面推进学校的信息化建设。



谢谢关注

电话: 13602861550

邮箱: 1667940811@qq.com