

# 区域教育大数据的设计与应用：以上海市闵行区为例

恽敏霞

上海市闵行区教育局

**摘要：**在大数据日益成为各领域改革和创新点的新时代，教育领域如何顺应和应对？本文具体阐述了“上海市闵行区”如何站在新的历史起点，从师生发展角度出发，设计和应用区域教育大数据的过程和方法。文章认为，教育大数据既是机遇也是挑战，既是技术也是思想。区域教育大数据的建设要有明确的规划设计，要有系统的平台支持，要有及时的多维分析，更要有面向未来挑战的决心和勇气。

**关键词：**数据；大数据；教育大数据

随着互联网、云计算、数据挖掘等技术的发展，“大数据”既作为一种技术和方法，也作为一种思想和观念，得到迅猛发展。可以说，大数据开启了一次重大的时代转型<sup>[1]</sup>。它以一种前所未有的方式，通过对海量数据进行分析，获得有巨大价值的产品和服务，或深刻的洞见<sup>[2]</sup>。大数据正在促进人类生活、工作、思维的大变革。

教育大数据是大数据的一个子集，相对于经济、医学等领域，教育大数据还处于起步阶段。它主要是指整个教育活动过程中所产生的以及根据教育需要采集到的，一切用于教育发展并可创造巨大潜在价值的数据集<sup>[3]</sup>。有学者认为，教育大数据有广义和狭义之分。广义的教育大数据泛指所有来源于日常教育活动中人类的行为数据，它具有层级性、时序性和情境性的特征；而狭义的教育大数据是指学习者行为数据，它主要来源于学生管理系统、在线学习平台和课程管理平台等<sup>[4]</sup>。大数据在教育领域作用的发挥，既要借助广义的教育大数据，更要利用狭义的教育大数据。

如何在当前信息技术发展的大环境下，合理利用教育大数据促进教育的深度

变革和持续发展，这成为一个具有重要实践意义的学术话题。本文以上海市闵行区为例，阐述如何从师生发展角度出发，探索区域教育大数据的可能价值和应用方法。

## 一、明确定位：区域教育大数据的价值追求

闵行区 1992 年由原上海县和闵行区合并而成，占地 371.68 平方公里，内设 14 个街镇，常住人口 253.43 万人。闵行形似一把钥匙，位于上海版图的中央。受经济发展、人口导入、生育高峰等因素影响，近年来闵行教育发展迅猛。闵行区基础教育共有法人单位 357 个，专任教师数 1.8 万，在籍学生总数 23 万。闵行区基础教育体量位居上海市第二位。

根据《上海市中长期教育改革和发展规划纲要(2010-2020)》、《上海市教育综合改革方案》和《闵行区教育改革与发展“十三五”规划》等文件精神，紧密结合闵行区经济社会发展定位，闵行区确立了“为了每一个孩子健康快乐成长”为核心理念，破瓶颈、补短板，坚持立德树人，深化课程与教学改革，践行教育优质化、信息化、国际化、个性化的发展路径，努力实现闵行教育内涵发展的五个现代化（以学生发展为中心的教育观念现代化，以核心素养为抓手的学习课程现代化，以“互联网+”为标志的教学技术现代化，以大数据分析为特征的教育管理现代化，以国际视野、综合素养为内涵的师资队伍现代化），由教育大区逐渐向教育强区转型发展。

面对新的发展形势和新的教育要求，如何进一步优化学校结构、完善设点布局，如何夯实教育基础、提升师生个性，如何优化教育管理、提供精准服务？一方面，要继续运用传统的、优质的经验和方法，持续开展实践，提高教育品质；另一方面，要大胆运用现代教育的理念和先进技术，找准关键部位，突破教育瓶颈问题。由此，闵行区确立了深度推进教育信息化的区域教育发展路径，明确了让教育信息化“助力管理、改进教学、支持评价、服务决策”的发展定位。要求在过去的基础上，强化全程化、全面化、全息化和价值化，重视教育“数据”的积累，逐渐从狭义的教育数据走向广义的教育大数据；探索教育“数据”的应用，让“大数据”为宏观决策、中观实践、微观选择等提供更科学、更及时的服务。借力区域教育大数据，最终形成基于五个教育云服务（教学云、教师云、学生云、家长云、学校云）的闵行现代化教育治理体系。

## 二、开发平台：区域教育大数据的基础建设

闵行区历年高度重视教育信息化工作，自2002年建成覆盖全区所有中小学的全光纤教育专网，完成了所有普通教室的多媒体改造，积极开展教师队伍建设、教学应用实践、教育管理推进等方面探索实践。经过几年努力，闵行区教育信息化区域应用工作取得显著成效。建成了覆盖全区的教育信息化网络，实现了“校校通”和“班班通”，具备了能够满足师生开展创新学习活动的信息化教学环境。覆盖全区的教育网络环境在上海市处于较先进水平，教育管理信息化在全区基本得到普及，闵行区成为教育部电子政务试点单位。

但是，相对来说，过去的教育信息化，关注硬件建设，且以条线式思维为主；偏重基础建设，且功能应用相对单一。面对新的发展形势，要根据区域教育发展实际，重新进行整体规划和设计。组织上，要有分有合、协同建设；时间上，要连接过去、指向未来；空间上，要明确中心、横向贯通；呈现上，要层次鲜明、多样表征；使用上，要基于实践、服务实践；管理上，要内外兼顾、浅表一致。

由此，闵行区确定，全面构建以学生成长数据为龙头的区域教育数据中心，完善教育基础数据库，从推进教育工作（行政与业务）和伴随采集数据的角度，重新建设或优化建设各类平台，定期将各类数据汇聚流入“区域教育数据中心”，形成以学生为“数据线索”的思维方式进行数据挖掘，形成基于区域教育大数据的教育咨询报告。

## 1. 区域教育大数据的框架和结构

按照“注重教育过程、强化数据共享”建设原则，首先设计和建设区域教育数据中心（具体框架见图 1），全面采集与汇聚各类数据，打破行政科室及业务部门之间的数据壁垒。第一，确定区教育数据中心负责确立基础数据标准化规范并作为对外提供基础数据的唯一来源。所有业务科室、部门以及教育局外部机构只能从区教育数据局中心通过接口获取各类数据；第二，区教育数据中心整体统筹，业务科室参与，确定业务数据标准与规范，着力汇聚、整合多套业务管理数据。第三，由区教育数据中心为各业务科室、部门、单位提供数据服务。不同科室、部门与单位根据业务与工作需求，向区教育数据中心提出数据需求，区教育数据中心将相应数据以数据接口方式对外提供服务。

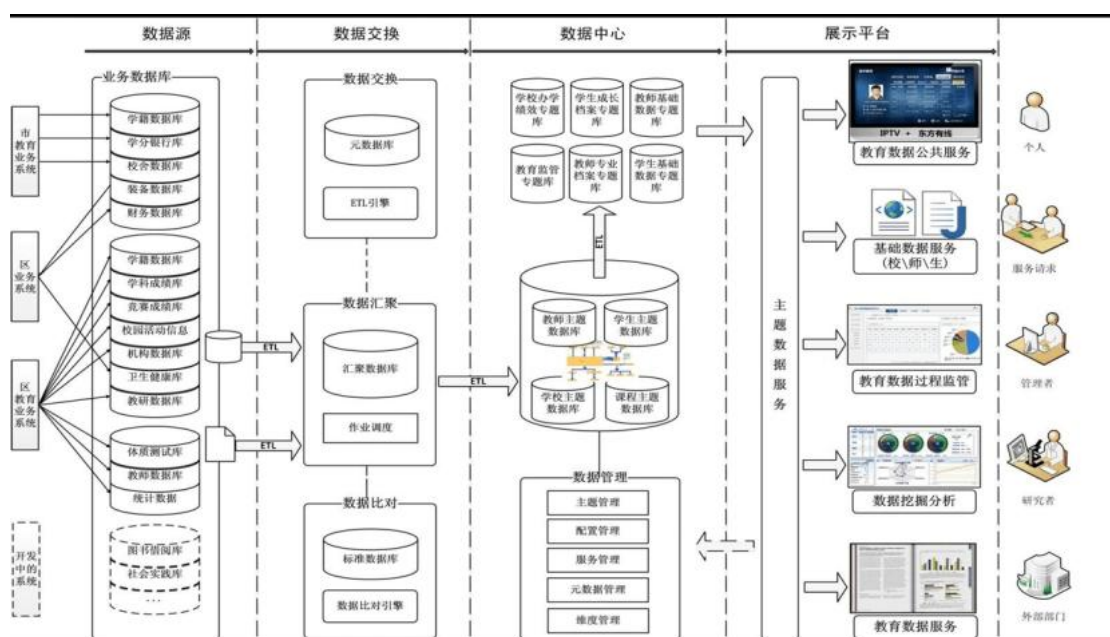


图 1 区教育数据中心框架

## 2. 区域教育大数据的信息化平台系统

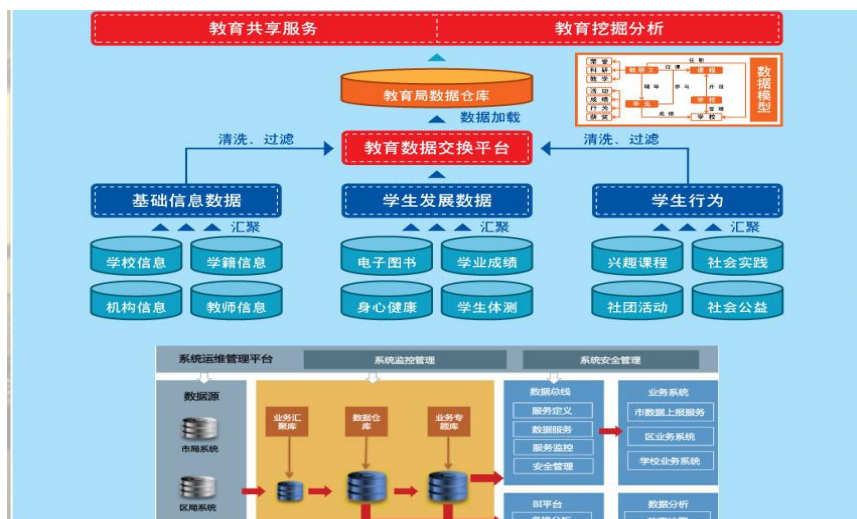
根据“数据中心”的规划和要求，设计和建设区域教育大数据的平台系统，明确各子平台的建设要求，由分管行政领导统筹协调，与业务工作对接，分工合作建设各个子平台（具体见表 1）。在各子平台建设的过程中，明确基础数据的规范要求和共享方法。

表 1 指向区域教育大数据建设的子平台（部分）

平台名称	主要数据	负责部门
学生电子成长档案系统	学生和学校基础数据、各类课程（包括学科课程）学习过程与结果数据、身心健康数据、社会公益与实践活数据等	信息化推进办
教师专业发展数字化支持系统	教师和学校基础数据、研修与成长数据、教学质量数据、专业荣誉数据等	区教育学院
学生体质健康检测平台	学生和学校基础数据、体质健康监测各类数据	区教育学院
学生校外活动管理平台	学生和学校基础数据、学生选修课程数据、学生所学课程过程与结果数据等	区青少年活动中心
学生电子书包及数字化教材应用系统	学生和学校基础数据、视频或教材阅读时间数据、资源选择数据、作业时长及质量数据、师生和生生交互数据等	区教育学院

### 3. 各类子平台的建设思路和要求

按照教育大数据的系统规划和“数据”要求，结合具体实践，各部门负责精细化建设每一个子平台，梳理数据关系，明确数据流通，确保各来源数据的真实性和可靠性。图 2 为学生电子成长档案系统的优化建设框架图。



在设计平台时，一方面，尽量从“数据”的规范性、科学性等角度进行规划，期望数据“可用、有用”；另一方面，从“数据”的真伪、效率等角度进行设计，尽可能伴随实践采集数据，确保数据的真实性、价值性。如学生校外学习因为自主性强，随机性大等原因，如果填报难辨真伪，则青少年活动中心将学生的校外学习从起点阶段（即选课）到结束（即学业质量，包括过程质量和结果质量）全流程架构在平台上，保证了全部学生、全部数据的真实性（如图3）。



**图3 学生校外学习数据流**

### 三、服务发展：区域教育大数据的应用领域

#### 1. 汇聚数据，初显区域教育大数据的特征

传统的教育数据因为单独建设，且多数通过业务流程进行管理和应用，因孤岛效应、数据差异等问题，很难实现面向服务对象的综合应用和持续应用。

闵行根据业务特点和教育大数据要求,从区域层面整体设计数据规范、质量标准、管理流程、分析方法等,这让教育数据的协同管理和综合应用成为可能。四年来,区教育数据中心已汇聚全区学生 6.1 亿个数据元,初步实现了狭义的教育大数据的应用价值(主要过程见图 4)。

## 2. 应用数据, 让教育大数据服务区域发展

数据的积累是为了应用,教育大数据的建设是为了更好地服务教育发展。闵

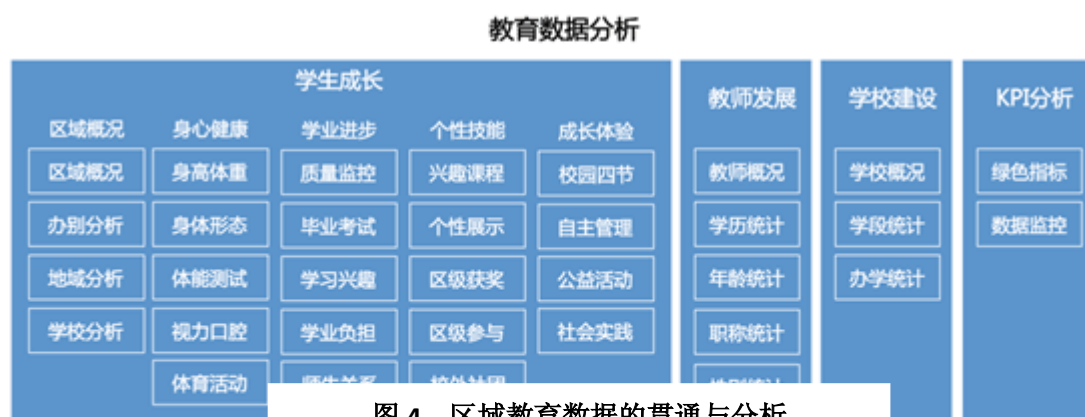


图 4 区域教育数据的贯通与分析



行教育大数据的建立,让其在“教师培训、学校绩效、学生发展(包括学业质量评价)”等方面有较好的应用。一方面,让区校教育政策的出台更精准、管理更完善。借助教育大数据,逐渐使行政化管理、经验性决策走向服务型管理、以数据为基础的决策<sup>[5]</sup>,整体提高教育管理与决策水平。如闵行每学年发布《闵行区学生电子成长档案数据分析白皮书》,白皮书从身心健康、学业进步、个性技能、成长体验四大维度对区域学生成长数据进行整体统计与分析,从数据角度反映全区学生学习及生活实际情况,探索全区教育教学中可能存在的趋势或潜在问题。同时,为全区不同学段、年级的学校和学生建立数据参照系。如发现异常,区教育局要及时出台相关政策

图 4 区域教育数据的贯通与分析



图 5 学生数据分析基本样态

另一方面，让区校对于教师、学生、学校的个性化干预、科学化培养成为可能。如闵行从“职业规划、师德修养、研修经历、教育实践”四个维度记录每一位教师的专业发展轨迹，构建了由教师研修网、课堂教学云录播系统、课程管理系统、学业质量分析系统、教师档案管理系统和教师个人空间等六大板块组成的教师专业发展支持系统，全面记录教师专业发展轨迹，迄今为止，已积累了 15174 名教师专业发展数据。基于这些数据的分析，区校两级层面皆可以对不同教师群体做出了专业特征评估，从而有利于区域开展更科学的人才决策和更有针对性的课程开发。图 6 为学校视角下，教师数据分析结果可视化呈现的部分样态；另外还有教师个体分析视角、区域整体分析视角、教师群组分析视角等。



图 6 学校教师数据分析基本样态

#### 四、综合应用：区域教育大数据的发展万问

教育大数据的应用要兼顾“数据收集、数据存储、数据处理、结果呈现”四个关键阶段的衔接性和专业性。因诸多原因，教育大数据未来的发展道路还很长。



从当前闵行的实践来看，主要要解决以下两个方面的问题。

### 1. 区域教育大数据的综合和单领域应用如何统筹协调

从闵行当前实践来看，各个领域已经能初步实现基于数据采集和分析，并基本实现常态化应用。但是平台之间的关系和数据之间的流动还需进一步明晰，综合和分项之间的原数据和应用边界要清晰。



图7 学校办学绩效督导评估系统

例如对于学校绩效督导评估系统来说，这不是原数据生成系统，而是综合应用数据的评估系统。但是在区域教育大数据整体设计过程中，偏重从原数据视角进行的平台系统的规划和设计。对于中间层的综合应用和全部系统的综合应用，究竟如何科学而便捷地实现，还需要进一步探索。

### 2. 区域教育大数据的采集和分析技术如何与时俱进

教育是一个复杂系统。无论对于整个教育生态还是一个学生的学习过程，都是一个动态的、受多因素影响的复杂过程。当前各类平台肯定不能采集到学生发展的“全数据”，尤其是对于教师成长和学生学习的过程性数据难以采集。另外，相对于数据采集来说，数据分析的专业技术要求更高，挑战更大。例如在学业质量分析中，我们使用了“聚类、决策树、IRT 潜能分析”等多个模型，但还未能实现这些分析技术的组合应用、自动化分析和可视化呈现。教育在革新，技术在发展，如何让两者能够互相顺应、与时俱进、共同发展，这需要多领域携手共研。

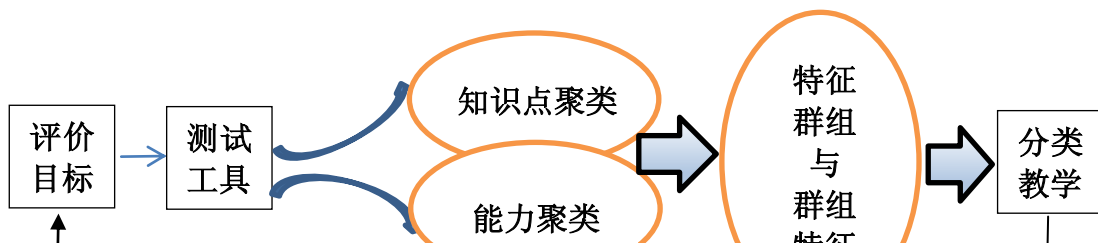


图 8 学业质量评价的聚类分析子模型

教育是一个逐步积累、不断进步的发展过程。关于闵行教育大数据的实践和研究，这是闵行在现代化教育主城区的发展变革过程中的必然选择和不懈追求。当前，因为理念、技术、方法、资源等多种原因，还处于教育大数据建设和应用的初级阶段。但几年的实践表明，教育大数据的建设过程就是思想转变的过程，教育大数据的探索性使用过程就是优化工作思路和改进教育工作行动的过程。面向未来，闵行将在教育大数据的两端继续加强研究，争取让区域教育大数据尽早焕发出催生教育变革的巨大力量。

#### 主要参考文献：

- [1]【英】维克托·迈尔-舍恩伯格、肯尼斯·库克耶著. 盛杨燕、周涛译. 大数据时代【M】. 杭州：浙江人民出版社.2013:9
- [2]【英】维克托·迈尔-舍恩伯格、肯尼斯·库克耶著. 盛杨燕、周涛译. 大数据时代【M】. 杭州：浙江人民出版社.2013:4
- [3]杨现民、唐斯斯、李冀红.发展教育大数据：内涵、价值和挑战【J】.现代远程教育研究.2016(1):50-61
- [4]徐鹏、王以宁、刘艳华、张海.大数据视角分析学习变革—— —美国《通过教育数据挖掘和学习分析促进教与学》报告解读及启示【J】.远程教育杂志.2013(6):11-17
- [5]胡弼成、王祖霖.“大数据”对教育的作用、挑战及教育变革趋势——大数据时代教育变革的最新研究进展综述【J】.现代大学教育.2015(4):98-104