

# 大数据时代区域教育数据网络 建设及关键问题探讨

杨现民<sup>1</sup>, 陈世超<sup>1</sup>, 唐斯斯<sup>2</sup>

(1.江苏师范大学 智慧教育研究中心, 江苏 徐州 221116; 2.国家信息中心, 北京 100045)

[摘要] 发展教育大数据是“十三五”期间各地教育信息化工作的重点之一。大数据在助力破解我国区域教育发展不均衡、教育决策粗放化、教育方式单调化、教育信息隐形化、教育就业盲目化、教育择校感性化等诸多难题上具有无限潜能。区域教育大数据的建设应以构建立体化的区域教育数据网络为基础,形成全区教育数据动态实时的采集、交换、共享、更新与创新应用的完备机制,为区域教育的全方位变革与创新提供强大动力来源和科学力量支撑。当前,我国各区县层面教育数据网络的构建常以教育数据中心为核心纽带进行整体架构,遵循四步走建设策略:“定标准,搭平台”、“聚数据,抓管理”、“聚数据,抓教学”、“用数据,促变革”。建设过程中涉及五大关键问题,分别是教育数据的采集机制、教育数据的融通共享、教育数据的开放运营、教育数据的安全保护以及教育数据的质量管理。最后,对区域层面和学校层面教育大数据的建设与发展提出了实施建议。

[关键词] 区域教育信息化; 智慧教育; 教育大数据; 教育数据网络; 关键问题

[中图分类号] G434 [文献标志码] A

[作者简介] 杨现民(1982—),男,河北邢台人。副教授,主要从事移动与泛在学习、智慧教育研究。E-mail: yangxianmin8888@163.com。

2015年是中国的教育大数据元年<sup>[1]</sup>。这一年促进教育大数据发展的政策文件正式出台(如《促进大数据发展行动纲要》明确提出建设“教育文化大数据”),确立了教育大数据在推动教育改革与发展方面的战略地位。中国教育大数据研究院、江苏省教育大数据科学与工程重点实验室等专业的教育大数据研究机构相继成立。市场上开始出现各种教育大数据产品,涵盖练考测评、课程教学、高考服务、家教匹配、学生管理等多项业务。地方政府、研究机构、学校以及行业企业力量的加入,正在逐步推动我国教育大数据事业步入发展的快车道。

当前,国内很多地区(北京、宁波、苏州等)都在积极探索和发展智慧教育。教育大数据汇聚存储了教育领域的信息资产,是发展智慧教育最重要的基础和保

障<sup>[2]</sup>,也是“十三五”期间各地教育信息化发展的重点方向。2015年3月19日,江苏省人民政府办公厅发布了关于推进智慧教育的实施意见,明确提出要“基于教育大数据,积极开展数据挖掘与学习分析,进一步优化教学策略、教学方式、教学过程和教学评价”。浙江省“十三五”教育信息化发展规划的主打方向也是智慧教育,同样明确要发展教育大数据,以支撑引领教育变革。本文首先介绍大数据在破解教育发展难题上的战略价值,接着探讨区域教育数据网络的构建思路,最后对教育数据网络建设涉及的关键问题进行初步分析,期望能对区域教育大数据的发展提供一定的借鉴和指导。

## 一、大数据助力破解区域教育发展难题

当前区域教育发展面临发展不均衡、教育决策粗

基金项目:江苏高校优势学科建设工程资助项目“江苏师范大学教育学省优势学科建设”(苏政办发[2014]37);江苏省“333”工程项目

放化、教育方式单调化、教育信息隐形化、教育就业盲目化、教育择校感性化等诸多难题。教育大数据建设不仅要持续、动态汇聚海量数据,更要实现数据基础上的教育“智慧”的创生。通过持续汇聚海量的教育数据,教育大数据能够为破解区域教育发展难题提供科学的数据分析与决策支持,为区域教育的全方位变革与创新提供强大动力来源和科学力量支撑,同时也有助于催生教育数据相关产业,带动区域经济发展。

### (一) 破解教育发展不均衡难题,实现教育普惠化

我国拥有世界上体量最大、结构最复杂的教育系统,区域教育发展极不均衡,严重阻碍了高质量教育公平的实现。教育信息化发展早期,由于技术条件的限制,难以获取全过程、全样本、细颗粒度的教育过程数据,容易出现供需偏差,也无法开展理想的个性化教育,让每个孩子都走向成功。大数据技术为破解区域教育均衡发展难题提供了可能,有望实现更加普惠和公平的教育。

我国三大教育基础数据库(学校信息库、教师信息库、学生信息库)的建立,以及统一规范的电子学籍制度的实施,将有效整合全国的教育数据资源,逐步形成规范的国家教育管理大数据。通过对这些数据资源的分析、建模和深度挖掘,一方面可以动态监管教师换岗、转岗以及学生学籍、转学、升学等轨迹,掌握区域教育发展现状;另一方面,可以结合经济发展、社会人口流动等,模拟未来区域教育运行状态,合理布局教育机构、调配师资、划拨教育经费等,进而实现高质量的区域教育资源均衡配置<sup>[3]</sup>。

### (二) 破解教育决策粗放化难题,提升决策科学化

传统的教育决策过分依赖于经验、直觉甚至流行趋势,而往往缺乏数据的支撑。随着大数据时代的到来,数据驱动的决策模式正在成为提高教育决策绩效的一个新视角,并逐步被各级教育管理部门以及各类学校管理者所接受<sup>[4]</sup>。大数据具有数据海量化、途径多元化、挖掘深度化、视角综合化等优势,有利于发现教育领域中同一类数据本身的特征、不同数据之间的联系、数据与外在影响因素的相互作用。它既可以全面透视宏观领域,又可以深刻剖析微观层面,易于构建更为系统化的教育发展模型,以推动区域教育政策制定与调整的科学化,而建立在大数据基础上的教育决策,还能增强民众对教育政策的理解和支持力,使得教育政策的制定和实施都更为科学化。此外,应用模拟仿真技术可以将各级各类教育机构长期积累的数据与其他行业数据(如医疗、交通、人口等)进行跨界

整合、交叉分析,动态模拟某项教育政策执行过程中可能出现的结果,以降低政策失败风险。此外,教育大数据还可用于检验教育政策的实效,比如上海市对小学一年级学生家长进行全样本调查,采集孩子作业情况、适应情况、学习压力等各项数据并进行分析,以检验上海小学教育零起点政策的执行效果<sup>[5]</sup>。

### (三) 破解教育方式单调化难题,助推教育个性化

客观来讲,我国多年的教育改革效果并不明显。信息技术变革课堂教学的零星案例虽然时有出现,但仍未形成主流,以教师为中心的单一、集中授课仍是当前各级各类学校课堂教学的主导模式。个性化教育、个性化发展、个性化评价、个性化学习……一直是广大教育工作者追求的目标,也是众多教育学者研究的热门议题。大数据时代的到来,有望融合教育“规模”与“质量”这一长久以来难以调和的矛盾体,让教师开展一对一的个性化教学,让学生享受多对一的个性化学习辅导。工业时代流水线式的同质化教育模式,将被私人定制式的个性化教育模式彻底取代。

当前,浙江、江苏等很多地方都在积极推进“走班选课”计划,多样化的选修课程为学生提供了更多选择。教育资源的极大丰富和多样化是开展个性化教育的基础,学习数据的全面、全程采集和深度分析是实现个性化教育的关键。大数据通过全程全面记录和分析学生的学习过程以及结果数据,能够精准识别每位学习者的学业优势、知识结构缺陷、兴趣偏好等,进而提供更有效的、更及时的学习指导,包括课程的选择、学习路径推荐、学习计划安排、学习内容强化等。个性化教育不仅指学生的个性化学习和教师的个性化教学,还包括教师的个性化发展。通过建立教师成长档案,动态汇聚各种教学、教研、课程、竞赛、论文、辅导等数据资源,可以为每位教师精准画像,一方面利于管理者实时掌握教师成长动态,更好地提供专业发展支持服务,另一方面利于教师更加清晰、准确地认识自己,进而明确个人发展方向,制定更有效的改进措施。

### (四) 破解教育信息隐形化难题,促进教育可量化

原来的教育信息都是隐形的,缺少搜集、汇聚、分析和公开,大数据促使教育变得可量化。云计算、物联网、大数据、可穿戴设备等现代信息技术的快速发展为海量教育数据的采集与深度应用提供了支撑,使得教育智慧化不仅只停留在表面,而是真正成为“有源”之水。这种“可量化”体现在教育的各个层面和环节,包括教学过程的可量化、校园管理的可量化、教育评价的可量化等多个方面。以课堂教学评价为例,大数

据可以精准记录每位学生的出勤率、回答与提问次数、每道练习题的完成时间与得分、走神次数等,然后以可视化的仪表盘方式一目了然地呈现给任课教师。课堂教学过程的量化能够辅助教师开展更加精细化的课堂管理和更加精准化的课堂教学。

#### (五)破解教育就业盲目化难题,指导择业合理化

教育资源信息不对称、人才供需矛盾等因素使得近年来大学生就业面临着更为严峻的形势。一个现实的问题是,很多大学生缺乏职业规划以及相关求职信息的准确判断,往往容易陷入盲目求职的“泥沼”,身心疲惫而又难以如愿。大数据能够帮助各级政府、学校以及企业为广大毕业生提供全面、合理、及时而又个性化的就业指导,帮助学生快速、理性选择就业方向 and 合适的职位。美国在利用大数据促进学生就业方面作了积极的尝试,为我国提供了宝贵经验。

美国劳工部基于多年的就业统计数据推出了“一站式就业服务”系统,该系统通过对全美历年行业薪资、教育培训、就业信息等数据的采集整合、分析处理,为公众提供了一个可以发现职业、薪资收入、教育培训、查找就业信息等功能的“一站式”服务网站。这种“一站式”服务背后正是利用了大数据技术,对教育、人力资源等多个领域的的数据信息进行关联分析后逐步生成。以其中的“薪资收入”为例,如果你是一个刚走入社会的大学生,可以先根据专业情况查找一下自己可能从事的行业,然后对比一下这些行业的薪资水平,从中找出一个薪资较高的行业。同时,你可能还需要对不同的城市生活成本进行比较,这样才能决定你要去哪个城市工作定居。

#### (六)破解教育择校感性化难题,推进选择理性化

由于教学资源分配不均带来的“择校”问题已经成为影响中国基础教育发展的一个难题。优质的教育资源少,分配不合理、不均衡,使“名校”效应更明显,学区房价格随之上涨,大数据无疑为解决教学资源分配不均的问题提供了新的途径和方法。择校难问题不是中国特有的,就全世界范围来看,多个国家普遍存在着这一共性问题。

近年来,随着教育数据开放步伐的加快,美国、英国等国家纷纷抓住机会,利用大数据对全国高等教育、基础教育资源等进行整合分析,搭建统一的教育资源平台,寻求从根本上解决择校难问题的途径。例如,英国推出了“寻找最好的学校”项目,包括“Find The Best University”和“Find The Best School”;美国推出了“大学导航”项目,以及“学校搜索”项目;荷兰则通过开放 Open Education Data API 数据接口,推出了

多款数据工具,如 schooltip.net、10000scholen.nl、scholenvinden.nl 等网站和 APP 应用。截至 2014 年 3 月,平均每个月有 5 万多名家长利用这些 APP 应用查找不同学校,为孩子找到最合适的学校<sup>⑥</sup>。这些项目和应用背后都隐约存在着大数据的身影,其本质就在于通过对全国上万个学校几十项指标海量数据的整合分析,最后以可视化的方式呈现给公众,帮助家长、学生根据需要找到最适合自己的学校,让择校不再是难题。

## 二、区域教育数据网络的建设

区域教育信息化发展早期,由于缺乏统筹意识和顶层规划设计,导致大量信息化系统的重复建设,直到现在很多地区的孤岛问题依旧未得到有效解决。此外,由于教育数据意识薄弱,很多宝贵的教育管理与教学数据未得到有效保存,甚至有些学校学生历年的考试成绩数据也残缺不全。在教育数据应用方面,很多区县教育管理者也尚未形成“数据决策意识”,教育数据的应用没有参与政策倡议、决定、执行、评估、复议的过程,没有形成“通过数据来指导行动,用数据提高教育决策效率和决策质量”的信息时代新型管理模式。大数据时代迫切需要建设国家和区域层面的教育大数据。

### (一)教育数据网络建设思路

国家教育大数据以管理类教育数据为核心,侧重从宏观层面为教育改革与发展提供更多、更好、更准的数据决策支撑。区域教育大数据的建设除了要服务教育管理与决策外,还应服务教学质量的提升,以及教师的专业发展。区域教育大数据的建设应以构建立体化的区域教育数据网络为核心,覆盖全区各类学校以及相关管理机构的各项数据业务,形成全区教育数据动态实时的采集、交换、共享与更新机制,为区域教育现代化事业的发展提供全面可靠的数据支撑服务。

区域教育大数据的建设是一项系统工程,应坚持如下基本原则。

(1)统筹规划,分步实施。区域教育大数据建设应结合不同区域的实际情况,依据教育整体发展战略与需求进行统一规划,“有重点、按计划、分步骤”逐步推进,保证数据采集与管理系统的完整性、灵活性与可扩展性。

(2)需求驱动,融合创新。教育大数据的建设应以教育需求为出发点,采集哪些数据、搭建哪些平台、制定哪些机制,都要围绕教育的现实需求以及未来发展需求来整体设计,利于推动区域教育信息化向融合创

新阶段发展。

(3)安全为基,递进升级。高度重视教育数据的安全与学生隐私保护,任何信息设备与系统的安装、集成与部署,都应进行严格的数据安全等级评估。同时,为满足教育业务发展的需要,应逐步升级完善教育大数据平台技术以及相应的管理机制。

(4)上下联通,开放扩展。区域教育大数据建设,一方面应考虑与上级教育数据系统的联通,满足上级管理部门对区教育数据采集与更新的需要;另一方面,应保证自身体系具备足够的灵活性和开放性,提供规范的数据开放接口,能够快速集成第三方提供的数据库应用。

随着国家教育“三通”工程的大力推进,全国各地的教育信息化服务能力和水平都有了显著提升,包括网络带宽、优质资源建设、应用系统部署等。调研发现,北京、上海、深圳等发达地区的教育信息化建设都在走向智慧教育综合服务平台(大平台)模式,即通过搭建大型教育云平台,部署各种应用服务系统,建立数据贯通与集成的有效机制,形成一体化、综合性、智能化的区域教育信息化服务体系,着力解决区域教育信息化发展过程中存在的系统孤岛、条块分隔、管理效率低下、优质资源匮乏等突出问题,促进区域教育现代化事业的快速、科学发展<sup>[7]</sup>。大平台建设正在逐渐成为各区信息化系统建设的主导模式,区大平台建设的通用技术架构如图1所示<sup>[8]</sup>。

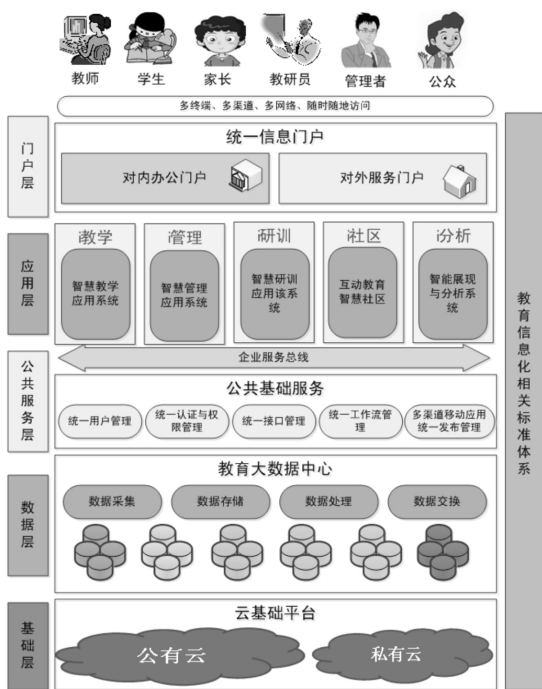


图1 区智慧教育综合服务平台体系架构  
区域教育数据网络的建设应积极纳入各地的教

育信息化“十三五”发展规划,作为其中的重点工程进行打造。建议的重点任务包括:(1)建设统一的区教育数据中心,汇聚来自每所学校产生的各类基础教育数据;(2)建立一套服务全区教育数据采集、处理、分析、管理与应用的规范化流程、标准与制度;(3)建立一支专业素质高、结构合理、分工明确的区域教育大数据管理队伍;(4)探索多种教育数据的创新应用模式,发挥数据在促进区域教育公平、教育资源均衡配置、个性化教学以及教师专业发展方面的重要价值;(5)建立区教育数据的合理开放机制,吸引更多社会机构参与教育数据的治理与相关工具服务的研发;(6)开展数据素养专题培训,提升教育管理者以及广大教师的数据意识与伦理道德、数据定位与采集能力、数据分析与解读能力以及数据反思与决策能力。

(二)教育数据网络体系架构

区教育数据网络的建设与大平台密切相关,数据产生于区、校部署的各应用系统,并在数据中心与各应用系统之间流动。调研发现,当前我国各区县层面教育数据网络常以教育数据中心为核心纽带进行整体架构(如图2所示)。

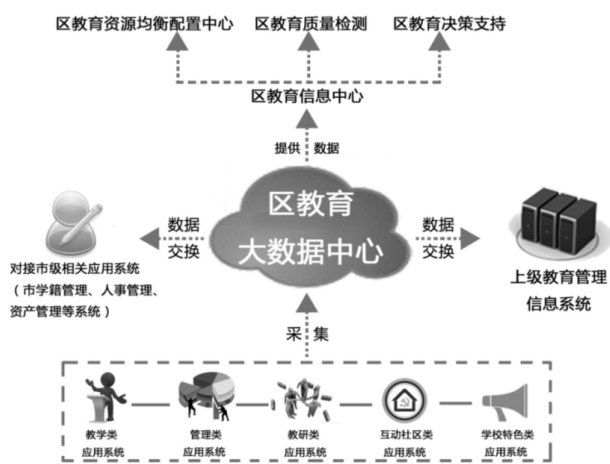


图2 区县层面教育数据网络架构

教育数据中心是区大平台的核心模块,是大规模数据存储和信息流通的中间节点,为全区提供教育信息共享服务,采集四大类云应用系统产生的基础数据以及来自各个学校特色化的应用系统产生的数据,在存放大量数据的同时有效地将数据管理起来,并提供数据访问的手段,为系统集成和各个系统之间的数据共享提供平台,保证数据的及时性、完整性和一致性<sup>[8]</sup>。

在数据互联互通方面,一方面区教育数据中心应与上一级的教育管理信息系统联通,根据数据更新要求适时传递、更新区教育管理信息到上级平台;另一

方面,还应与市级其他相关业务系统联通,比如与市学籍管理系统、设备资产系统、人事系统等实现相关信息的互换共享。区教育信息中心是区大平台的主管单位,统筹整个区大平台的建设、管理、应用与维护,负责全区教育数据存储与安全管理,同时辅助教育局相关部门编制区教育事业发展报告,开展区教育信息化、教育现代化发展水平评估。

### (三)教育数据网络建设步骤

严格来说,区域教育大数据的发展并没有统一的建设和实施路径。然而,我国教育大数据的发展整体上处于起步阶段,缺少可供参考的实践案例和经验。为了进一步促进国内区域教育大数据的发展,接下来将对区域教育大数据的实施路径进行概要设计,期望能起到一定的指导作用。区域教育大数据的发展是一个从低级向高级发展的过程,其建设与应用推进大体可分为四个阶段(如图3所示)。这四个阶段发展重点不同,但又不是决然分开,发展任务之间存在一定的交叉和重叠。

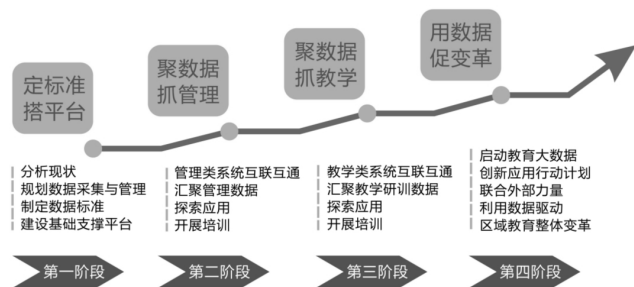


图3 区域教育大数据实施路径

**第一阶段:定标准,搭平台。**全面梳理分析全区现有各种教育信息化应用系统,结合区域教育发展战略与实际需求,规划教育数据的采集范围、技术与机制。综合全国、省、地市现有教育信息化标准,制定区域教育大数据建设标准。应用云计算技术,建设区域教育数据中心基础支撑平台。

**第二阶段:聚数据,抓管理。**打通数据中心与各个管理应用系统之间的数据传输渠道,重点实现各类教育管理数据的规范化采集、存储、共享与分析。实现与国家、省等上级管理信息系统的互联互通。积极探索基于数据的教育管理模式变革,研发教育决策支持、教育质量检测、教育发展评估等数据应用系统。开展面向教育管理人员的数据素养提升专题培训。

**第三阶段:聚数据,抓教学。**打通数据中心与各教学、教研应用系统之间的数据传输渠道,重点汇聚教与学以及教研培训过程中的各种有效数据。应用教育数据挖掘、学习分析、可视化等关键技术,研发促进学

生个性化学习、教师个性化发展的各种应用系统与智能工具。鼓励教师探索大数据在促进教学方面的创新应用模式,积累优秀案例并分享。开展面向教师、教研员的数据素养提升专题培训。

**第四阶段:用数据,促变革。**根据前几个阶段汇聚的海量教育数据,以及及管理、教学方面积累的数据应用经验,启动教育大数据创新应用行动计划。引入外部高水平的科研团队和智囊团队,全面探索、推广大数据在促进区域教育发展及变革方面的应用模式与实践案例,形成全区利用数据开展教育管理、决策与教学活动的文化,实现数据驱动下的区域教育成功变革。

## 三、建设过程中关键问题探讨

### (一)教育数据的采集机制

教育数据的全面、实时、持续采集是建设教育大数据的基础性和先导性工作。建立规范化的、动态连续的区域教育数据采集机制,有助于全面跟踪所有在校生的学习情况以及毕业生的工作情况,进行更加客观、更加全面的区域教育质量评价,并依据评估结果针对性地优化、调整区域教育服务体系,包括专业设置、师资配备、课程开设、实践教学等,有效解决当前教育系统与社会需求脱节这一现实问题<sup>[4]</sup>。

区域教育大数据建设需要综合应用多种数据采集技术,共包括四类(物联感知类技术、视频录制类技术、平台采集类技术及图像识别类技术)、十三种常见技术。每种技术采集的数据范围和重点有所不同,图4展示了教育数据采集的技术图谱。

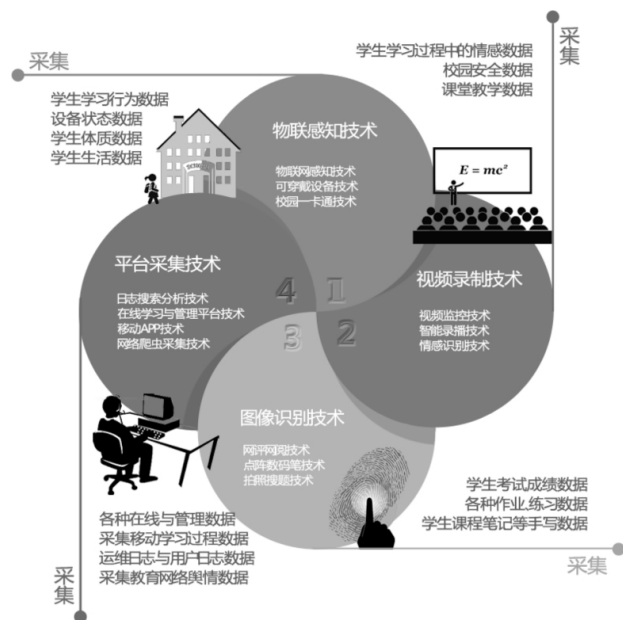


图4 教育大数据采集技术图谱

物联网感知技术主要用于采集设备状态数据和学生体质数据;视频监控技术主要用于采集校园安全数据;智能录播技术主要用于采集课堂教学数据;网评网阅技术主要用于采集学生考试成绩数据;点阵数码笔技术主要用于采集各种作业、练习、考试数据;拍照搜题技术主要用于采集学生作业练习数据;情感识别技术主要用于采集学生学习过程中的情感数据;日志搜索分析技术主要用于采集运维日志与用户日志数据;在线学习与管理平台技术主要用于采集各种在线学习与管理数据;移动APP技术主要用于采集各种移动学习过程数据;可穿戴技术主要用于采集个体生理数据与学习行为数据;校园一卡通技术主要用于采集各种校园生活数据;网络爬虫技术主要用于采集教育舆情数据。

为了保证高质量教育数据的可持续性采集,在实践过程中需要注意如下事项。(1)采集需要提前规划设计,涉及内容包括:数据采集的范围、使用的数据采集技术、数据采集环境的部署、数据采集质量的保障措施、采集数据的应用目的和场景、数据的存储方案、数据的更新机制、数据的交换标准等。(2)采集需要有清晰的边界,而非盲目采集任何教育活动数据,采集的数据范围取决于数据的应用目的。(3)采集要保持连续性和规范性,秉持“持续创造价值,规范提升价值”的理念。(4)在保证数据有效性的基础上,数据粒度应尽可能细小,以便从中挖掘更多潜在价值。(5)采集要符合伦理道德,数据产生主体应当有一定的知情权和选择权。

## (二)教育数据的融通共享

教育数据在不同应用系统之间的无缝流转与整合共享,是发挥教育大数据核心价值的基础和关键。应用云计算技术建设区域教育信息化综合服务平台(以下简称“大平台”),是我国区域教育信息化推进过程中的必然选择。目前,国内大多数区域教育信息化平台已经具备了单点登录功能,但这仅仅实现了认证端数据的共享,是区域教育数据整合的基础和开始。各应用系统之间基础数据的透明流转和无缝对接,仍是当前各地智慧教育综合服务平台建设的难点。

为了促进全区教育信息的无缝流转与数据共享,需要对区域内现有的应用系统和新开发的应用系统进行底层数据的整合。大平台完美集成的关键在于底层数据的流转共享,一个可行的方案是采用星形网络结构建设区教育数据中心。

区教育数据中心是所有大平台应用数据存储和信息流通的中心节点,为各系统之间的集成和数据互

换提供标准化的传输接口和通道,同时保证各项教育数据采集、存储与传输的完整性、及时性和一致性。区教育数据中心主要承担两项职责:一是汇聚来自所有数据源的规定数据,同时保持同步更新,支持教育数据在不同系统之间的无缝流通;二是作为综合性的教育数据仓库,为上层各项智能应用服务提供挖掘、分析所用数据。

统一数据标准的制定是区教育数据中心建设的基础性工作。现实情况是很多地方已经部署了来自不同厂商的教育应用系统,大平台的建设更多是对现有系统的整合。因此,比较适合大多数地区的做法是建立教育数据交换标准,明确规定基础性教育数据的名称、格式、字段等,同时提供统一的数据交换接口标准,以便各应用系统可以按照标准要求开发Web Service接口,将规定的各项数据传递到教育数据中心。

教育数据中心可以根据业务属性设置一系列主题数据库,比如学校信息库、学籍库、教职工基础信息库、学生学习与成长信息库、教师教学及成长信息库、试题试卷库、课程库、成绩库、教材库、设备资产库、公文库、科研库、财务库、资源库等<sup>[7]</sup>。主题库呈层级结构,每个主题库又可细分为若干子主题库,通过关联数据表保存来自相关应用系统的教育数据。教育数据中心的核心理功能包括:元数据管理、主题管理、数据采集、数据交换、数据抽取与同步、角色与权限管理、日志管理、数据安全管理和系统设置。

## (三)教育数据的开放运营

随着公共数据开放战略的推广和实施,越来越多的公共部门加入到开放数据的大军当中。教育数据作为一种特殊的公共数据,也需要在合理的范围内以合理的方式对外开放。支撑教育大数据的信息化处理、数据储存、后期维护等工作所需的经费投入,也会给教育主管部门、学校、在线教育培育机构等带来较大的经济压力。选择一种合适的教育大数据运营模式就显得极为重要,它是保证教育大数据良性发展的重要前提,也是提升教育数据服务水平的重要保障。当前,主要存在四种可能的教育数据运营模式。

### 1. 教育主管机构的教育数据运营模式

(1)非营利的教育数据。“自由和免费”是公共数据开放的原则,教育数据开放不仅是教育主管机构加快信息公开的有力推手,也是促进教育公共服务发展模式创新和推动教育改革的难得机遇。教育主管机构在保证个人隐私和安全的基础上可以适度、适时免费向社会开放部分教育数据。

(2)教育数据的商业再利用。教育主管机构针对一些价值较高的教育数据,利用商业模式,对开放数据进行再度开发利用,进而为用户提供增值服务。免费开放的教育元数据有时候对于公众而言参考价值不大,若教育主管机构运用其所掌握的技术优势和跨部门资源共享优势等,对教育数据进行二次加工,可以为公众提供更有价值的服务。因此,教育主管机构可发展具有特色的定制服务和产品,通过B2B或B2C的模式为用户提供增值服务,并收取相关的合理费用。此外,还可创建“数据再利用应用池”机制,汇集社会各界对教育开放数据的开发应用(付费或免费),从而调动公众参与教育的积极性,鼓励产业创新<sup>[9]</sup>。

## 2. 企业的教育数据运营模式

国内更多关注的是互联网在线教育行业的运营模式,如B2C、B2B2C等多种模式。企业的教育数据运营模式强调的是如何利用教育数据这种新资源去盈利、去提升竞争力和抢占市场。过去几年,在线教育领域发展的一种趋势是行业巨头和投资人持续加大投入,拓展业务;另一种趋势则是众多创业者对在线教育进行细分,努力寻找新的“蓝海”。

两种发展方向催生出两大教育大数据营运模式,一方面是以淘宝、百度等巨头为代表的平台化运营,对平台上的各种教育机构所提供的教育数据进行整合,提供更加体系化的、全流程的教育服务;另一方面是以粉笔网为代表的精细化运营,对在线课堂、学生管理、题库等教育数据进行深度挖掘,提供更加有针对性、个性化的教育定制服务。

但是,我们也要清醒地看到,企业的教育数据运营模式存在着较大的政策风险和市场风险。以美国inBloom公司的“被倒闭”为例,其失败的很大一部分原因在于其商业模式的失败,它的经营模式简言之就是一条“收集数据—兜售数据”的数据渠道,从不同的学校收集学生数据,经过加工处理使得数据标准化,将数据分享给其他需求者。此外,对数据隐私和数据安全两大“短板”缺乏必要保障措施,忽略了通过提供用户端软件与用户进行“交互”的过程,也是其倒闭的重要原因。与inBloom不同,成立于2012年的Clever公司不存储任何教育数据,而是通过独立的API让不同应用系统产生的海量学生数据更容易被其他教育软件开发商获取和利用。迄今已有2万多所学校与Clever签约,且未被公众质疑。

## 3. 公私合营(PPP)的教育数据运营模式

(1)特许运营。随着教育数据的快速增长,其体量的巨大化、结构的复杂化、处理速度的快速化等专业

技术要求越来越高,一般的政府或学校的信息化支撑部门,已经不能完全、高效地满足教育部门对教育数据的应用和管理需求。这时,政府部门就需要借助企业等市场力量,采用第三方委托方式,与更有技术优势、内容优势的企业进行合作,使教育数据运营管理效益最大化。

(2)数据“捐赠”。企业通过参与教育信息化项目建设、运维、教育培训等,也会创造和收集一些数据,并对其挖掘分析形成新数据,这些数据将会作为企业的重要资产,而政府公共部门可能接触不到相同规模、质量或频度的此类数据。为进一步建立资源共享、互惠共赢的合作关系,政府公共部门会和企业进行合作,企业向政府“捐赠”数据,政府获取这些数据资产的使用权后得以掌握更加全面的数据量。在教育领域,随着互联网在线教育规模的扩大和领域的细分,企业掌握的教育数据资源会越来越多、价值会越来越高,特别是帮助教育主管部门制定教育决策、改善教育环境等方面,企业可以在保障用户隐私的前提下将教育数据“赠予”教育主管部门。这种“捐赠”方式的公私合营运作模式可以有效弥补单纯以盈利为目的的企业和非营利的政府部门间的合作空白。

## 4. 公众参与的“众筹”运营模式

“众筹”是一种新兴的商业运营模式,即大众筹资或群众筹资,以公众参与为核心的运营模式<sup>[10]</sup>。随着信息时代公众参与社会治理的积极性不断升级,众筹模式已经开始被应用到公共服务和社会治理领域,公共服务模式也从政府主导的“G2C”模式向公众主导的“C2G”模式发生转变。在教育领域,众筹运营模式的应用前景十分广阔。例如美国的公立学校一向以教育经费少、师资缺乏、学生素质差著称,DonersChoose.org对此搭建了一个类似公益网站的平台,允许公立学校的教师上传改善自己学校条件的项目,用户通过平台自由选择资助对象<sup>[11]</sup>。

教育作为公共服务的重要组成部分,在教育大数据运营中可以尝试引入以众筹为代表的“众包”服务模式。教育是覆盖全社会不同主体、不同领域、不同行业的公益性活动,需要调动起各类主体参与建设的积极性。教育主管部门、学校及其他教育培训机构等只是教育产业链中的一小部分,随着全民教育、终身教育理念的普及,以家长、学生等普遍公众为主的全民才是教育产业链中的真正主体,更是教育大数据的主要来源。如何发挥普通公众的“长尾”效应,充分发挥公众在教育大数据中的作用,让公众变为教育大数据的“数据源”,形成数据“众筹”服务模式,是未来教育

大数据运营的创新点。

### (四)教育数据的安全保护

在传统应用场景中,数据由学校等部门局部收集,在校内应用,数据所有权的问题并不突出。在大数据时代,数据要更多进行跨界整合、外部应用的商业模式创新,这其中就涉及更多数据所有权的话题。数据安全已成为影响大数据的顺利发展的关键因素,区域教育数据网络的建设也不例外,亟须加强对教育数据的安全防护。建议采取如下措施。

#### 1. 建立完备的教育数据安全监管体系

组建由教育主管部门、数据服务提供商、学校和第三方测评机构协同管理的教育大数据安全监管体系,制定数据服务安全标准,定期实施教育大数据安全评估。依照国标计算机信息系统防护要求,结合大数据治理,国家组织相关业务人员、技术人员和企业,根据具体情况汇总并制定符合大数据建设的安全管理目标和范围,制定符合教育系统实际的安全管理策略。各级教育主管部门按照总体安全策略和业务应用需求,建立教育数据安全管理机构,制定数据安全管理的规程和制度。相关人员必须严格按照此管理制度和安全策略执行。

#### 2. 加强教育数据安全队伍建设

在各教学单位设置专职数据安全管理人员,让具有安全管理工作权限和能力的人员担当,坚持“权限分散、不得交叉覆盖”的原则,系统管理员、网络管理员、数据库管理员不得相互兼任。对参与人员进行背景和资质审查以及技能考核,同时签署数据保密协议。对离岗人员限制其所有访问权限,收回相关证件、设备及访问控制标记等。定期对各岗位人员进行不同侧重的安全认知和安全技能培训和考核。通过对管理类体系测评办法的考察,以及参考国内安全标准中关于对管理层面的考核,将安全管理机构、人员岗位设置、岗位职责、系统运营维护、应急事件管理纳入管理类考核指标。

#### 3. 加强系统运维管理和定期安全评估

对各设备的操作应由授权的管理员进行,按操作规程实现设备的启动/停止、加电/断电等操作,维护设备的运行环境及配置和服务设定,加强日志文件管理和监控管理,加强配置文件管理,定期对系统安全性进行有效性评估和检查,及时发现系统的新增缺陷和漏洞。

对于教育数据的日常运行和管理,各单位应按审查和批准的用户分类清单建立用户和分配权限,说明各用户权限、责任人员和授权记录,定期检查特权用

户和重要业务用户的实际分配权限是否与清单相符,对关键业务用户开启审计功能。

在日常运行中,应派专人定期检查安全策略执行情况,并生成相应报告。对各个岗位定期检查操作规程和管理程序的执行情况,确保遵从安全管理小组制定的安全策略。教育数据管理人员应防止违法行为,对设计、操作、使用和管理应规避法律法规禁区,保护组织机构的数据信息和个人信息隐私,有措施防止对信息处理设备的滥用,避免危害机构和社会利益。

采用多层面、多角度的系统分析方法,由用户和专家对资产、威胁和脆弱性等方面进行综合性评估,建议处理和减缓风险的措施,形成风险评估报告。根据风险评估结果,结合有关安全保密防护要求,决定各子系统的安全控制措施,对各控制措施进行综合分析,得出紧迫性、优先级、投资比重等评价。

#### 4. 不断升级教育数据安全防护技术

根据教育数据价值范畴以及隐私层次的不同,设定数据的不同安全等级,明确治理目标,分析安全盲点,降低数据泄露风险<sup>[12]</sup>。对于重要数据的传输和存储应采用严格的加密技术,保证数据安全。

建立全网覆盖、集中控制的病毒防护体系,对检测和截获的各种高风险病毒进行技术分析处理,提供相应的分析报表和总结汇报。按照相关规定对使用的密码算法和密钥进行管理,对信息系统中以密码为基础的安全机制分等级管理,制定定期更换密码策略。

针对教育系统运行特点,各单位应制定相应的备份和恢复策略,形成备份策略文档,明确说明需定期备份的重要业务信息、系统数据及软件等内容和备份周期,确定重要业务信息的保存期及其需要保存的归档副本的保存期。定期检查备份介质,定期测试恢复程序,指定专人负责数据备份和恢复,并采用热备份方式保存数据,定期执行数据增量备份和应用环境的离线备份。对关键区域应建立全冗余硬件部署,包括网络设备、存储设备以及服务器。

发生入侵、病毒破坏、非法访问等安全事件后,根据事件对系统的破坏程度、造成的社会影响及涉及的范围确定系统安全事件处置等级,建立数据安全事件分级响应、处置制度。建立切实可行的应急处理流程,指定专人负责应急处理计划和实施恢复计划管理工作,各级信息安全管理应负责具体落实、检查和验证应急处理计划和灾难恢复计划,保证计划能够有效执行。

### (五)教育数据的质量管理

数据质量是数据分析结论有效性和准确性的基



础,也是最重要的前提和保障。如果采集的教育数据质量有问题或者根本就是错误的,那最终得出的任何结论都将“误导”教育实践。教育数据质量的评判标准包括规范性、完整性、准确性、唯一性、关联性和一致性。一般来说,影响教育信息系统数据质量的因素主要有以下几种<sup>[13]</sup>:数据系统缺乏兼容性、缺乏标准化的数据定义、数据本身的可用性不强、数据提供者对同一数据项理解不一致、数据缺乏时间持续性、数据录入存在错误、数据更新不及时。为了保证教育数据质量,建议采取如下措施。

#### 1. 树立数据质量管理意识

大数据时代教育数据正在成为学校一种无形的战略资产,在支持创新人才培养方面发挥着不可估量的价值。数据质量是教育大数据的生命线,质量低下、甚至错误的数据必然导致错误的分析结果,极有可能误导教育决策、教育管理以及教学活动的开展。因此,所有教育数据的管理者和使用者都应牢固树立数据质量意识,切实把教育数据当作资产对待。不仅考虑从技术层面着手加强教育数据在采集、存储、处理、分析、应用等各个环节的质量,还应着重从管理层面制定完备的教育数据质量管理制度和规范,包括数据审查流程、岗位职责确定等。

#### 2. 制定数据质量管理规范

教育数据质量管理规范既包括教育数据的技术处理标准,也包括教育数据的管理业务标准。一方面,应基于现有教育信息化标准制定专门的教育数据技术处理规范,包括数据采集、数据录入、数据传输、加密处理等具体要求;另一方面,应结合机构现有管理规章制度,制定完备的数据管理流程和业务规程。有条件的机构建议成立专门的部门负责教育数据的质量管理工作,设立专职或兼职的数据质量管理岗位,定期监控各应用系统教育数据质量状态,编写教育数据质量问题报告<sup>[14]</sup>。

#### 3. 运用数据质量检查工具

数据质量管理是一项极其重要而又烦琐的工作。为了实现教育数据质量的高效管理,需要结合系统实际开发专业的数据质量智能检查工具,以辅助管理员快速检测教育数据的一致性、重复性、规范性和完整性。对于区域智慧教育综合服务平台而言,建议单独

部署教育质量管理体系,以统一管理和维护来自各应用系统以及教育数据中心的数据质量。

## 四、总结与建议

教育大数据是区域教育事业发展的重要资产,也是推动区域教育科学创新、精准变革的重要力量。为了促进区域教育大数据健康、有序发展,提出如下建议:(1)鼓励有条件的区县,优先采用国内云服务厂商提供的技术,搭建区教育数据中心;(2)区教育数据中心的建设不是“从头再来”,应有效整合已有硬件、软件资源,保护已有投资;(3)共性的教育业务系统应统一部署到区大平台上,同时遵循统一的数据及互操作标准,与区教育数据中心之间互联互通;(4)加强教学大数据的规范采集、存储、管理与分析应用,构建各学科知识模型,对学生的学习过程与结果进行诊断、预测与有效干预,精准提升教学质量;(5)全面评估教育数据开放的风险与应对措施,将部分教育数据服务委托给有资质的社会机构进行专业化运作,切实提高教育机构的专业服务能力。

中小学校是教育数据网络的“神经末梢”,原则上不建议单独建设校级数据中心。除非一些高带宽要求(如高清视频互动系统)或比较特色化的业务系统,可以部署到学校机房外,其他任何业务系统都应优先使用区综合服务平台统一提供的服务。学校无须购置数量庞大的、昂贵的服务器设备,只要将校内各种计算终端与区教育云连接,通过浏览器、APP等即可访问各种教育应用系统,使用方便,成本低廉,安全性高。

针对学校层面教育大数据的建设与应用,提出如下建议:(1)建设高速校园无线网络,支持BYOD运动,让教师、学生可以通过多种终端,方便而又安全地获取云教育平台提供的各种服务,持续累积学与教数据;(2)鼓励教师利用教育数据探索新型教学模式,不断优化教学,归纳提炼有推广价值的驱动教学的案例;(3)加强学生网络与终端使用安全方面的专题培训,从“源头”降低学生隐私数据泄露风险;(4)不断提升学校管理者与教师的数据素养,形成一种使用数据进行管理决策与教学决策的氛围;(5)采购学校信息化相关产品时,应全面了解产品对教育数据的采集范围、存储与应用方式等,将教育数据安全作为重要考核指标。

### [参考文献]

- [1] 杨现民,唐斯斯,李冀红. 教育大数据的技术体系框架与发展趋势——“教育大数据研究与实践专栏”之整体框架篇[J]. 现代教育技术,2016,26(1):5-12.
- [2] 柯清超. 大数据与智慧教育[J]. 中国教育信息化,2013(24):8-10.

- [3] 刘雍潜,杨现民. 大数据时代区域教育均衡发展新思路[J]. 电化教育研究,2014(5):11-14.
- [4] 杨现民,王榴卉,唐斯斯.教育大数据的应用模式与政策建议[J].电化教育研究,2015(9):54-61.
- [5] 虹口政协. “大数据”呈现上海小学教育“零起点”政策实效 [EB/OL]. (2015-03-04)[2015-12-26].<http://www.hkqzx.gov.cn/StaticArticle/150304/134848974.shtml>.
- [6] For kids' sake, connecting parents to Dutch schools with open data[EB/OL]. (2014-02-17)[2015-10-11]. <http://www.epsiplatform.eu/content/kids%E2%80%99sake-connecting-parents-dutch-schools-open-data-0#sthash.7McGcjFw.dpuf>.
- [7] 杨现民. 区域智慧教育综合服务平台建设及关键问题探讨[J]. 现代远程教育研究,2015(1):72-81.
- [8] 许鑫,苏新宁,吴乃冈. 高校共享数据中心平台的设计与实现[J]. 现代图书情报技术, 2005,21(6):48-53.
- [9] 法国探索公共数据开放良性发展新模式 [EB/OL]. (2014-08-05)[2016-06-27].[http://cyw.cena.com.cn/2014-08/05/content\\_236397.htm](http://cyw.cena.com.cn/2014-08/05/content_236397.htm).
- [10] 杨东,苏伦嘎. 股权众筹平台的运营模式及风险防范[J]. 国家检察官学院学报,2014,22(4):157-168.
- [11] 翟葆光.用科技撬动臃肿传统的政府职能[EB/OL].(2013-03-27)[2016-06-27].<http://www.pingwest.com/crowdfunding-public-projects/>.
- [12] 代欣. “大数据”环境下高校信息安全探索[J]. 信息技术与信息化,2015(3):12-13.
- [13] 郑燕林,柳海民. 大数据在美国教育评价中的应用路径分析[J]. 中国电化教育,2015(7):25-31.
- [14] 宫剑,高亮. 上海财经大学:提升数据质量之道[J]. 中国教育网络,2014(1):71-73.

### A Study on Construction and Key Issues of Regional Education Data Network in the Era of Big Data

YANG Xianmin<sup>1</sup>, CHEN Shichao<sup>1</sup>, TANG Sisi<sup>2</sup>

(1.Research Center of Smart Education, Jiangsu Normal University, Xuzhou Jiangsu 221116;

2.National Information Center, Beijing 100045)

**[Abstract]** The construction and application of big data in education is one of main tasks of educational informatization in various regions during the 13th five-year plan. Big data has infinite potentials in solving many problems in regional education, such as the imbalance of educational development, the extensive educational policy making, the monotony of education modes, implicit educational information, the blindness of employment and the sensibility of school choices etc. The construction of big data in regional education should be based on constructing three-dimensional regional education data network, and then a complete mechanism that contributes to real-time acquisition, exchange, sharing, updating, application and innovation of education data in the whole area can be formed to support the reform and innovation of regional education. At present, the construction of education data network in each district or county usually centers on educational data and follows a four-step strategy: setting standards to build platform; collecting data to stress management; gathering data to stress teaching; using data to promote reform. Five key issues involved in the process of construction are: acquisition mechanism of educational data, combination and sharing of educational data, opening operation of educational data, security of educational data and quality management of educational data. Some implementing suggestions are put forward in this paper about the construction and development of big data in education in regions and schools.

**[Keywords]** Regional Education Informatization; Smart Education; Big Data in Education; Education Data Network; Key Issues