

美国高等教育综合数据系统： 内涵、运行、特点与启示*

王兴宇

摘要：教育数据系统建设在高等教育管理信息化过程中扮演着重要角色。美国高等教育综合数据系统（IPEDS）作为高等教育的基础数据库在积累基本教育资料、掌握教育概况、提高管理效率等方面发挥了巨大作用。文章从内涵阐释入手，通过分析该系统的数据来源、数据采集和数据利用等运行机制要素，发现整个数据系统呈现出整体规划、三级管理、赋权增能等特点。我国教育数据系统建设可以从顶层设计、技术引进、分析应用和文化培育等方面借鉴美国高等教育综合数据系统的经验。

关键词：美国；高等教育；综合数据系统；数据

教育数据系统建设在高等教育管理信息化的过程中扮演着重要角色。它是教育领域综合改革的科学力量，对于系统收集各类教育数据，科学反映国家和地方教育现状，促进教育决策科学化，推进教育治理现代化，提升教育服务个性化，增强公众对高等教育的了解具有非常重要的意义。在联邦政府的支持下，美国国家教育数据系统发展迅速，处于世界领先地位。其中，高等教育综合数据系统（the Integrated Postsecondary Education Data System, IPEDS）作为美国高等教育的基础数据资源库在积累基本教育资料、掌握教育概况、提高管理效率等方面发挥了巨大作用。

一、美国高等教育综合数据系统的内涵

美国高等教育综合数据系统（the Integrated Postsecondary Education Data System, IPEDS）是美国国家教育统计中心（National Center for Education Statistics, NCES）授权国家高等教育合作社（National Postsecondary Education Cooperative, NPEC）研究和开发的教育数据项目，是美国高等教育领域的核心数据采集计划和信息系统。1965年修订后的高等教

育法案明确提出，所有参与联邦学生资助计划的大学和学院有责任向教育部门提交申请和注册学生数、学位颁发和证书获得情况、毕业率、教职员工、财政情况、学杂费以及学生资助等教育数据，并对此作出了强制性规定，所以IPEDS所有调查部分的响应率几乎都接近100%。^[1]目前，有超过7300个高等教育机构（涉及50个州和哥伦比亚特区）依法向IPEDS提交相关的教育数据资料，IPEDS已成为美国最全面和最完整的高等教育数据来源之一。^[2]

IPEDS的数据是完全公开和免费的。在每年三次的收集工作结束后，IPEDS会有六个月的时间对数据进行处理，通常前三个月会对数据进行全面审查、编辑和估算，并发布临时性报告，接下来的三个月教育统计信息管理人员会对数据进行深加工、再修订，并发布多种教育主题的正式性报告。^[3]同时，IPEDS还会根据每年数据分析结果发布大量相关的数据产品，例如IPEDS招生手册、毕业率手册、教育统计摘要、少数族裔学生群体的教育现状和发展趋势等。从其内容看，这些教育数据产品将原来分散的教育数据进行更为结构化的处理，使得教育统计信息转化为知识内涵更为丰富的

收稿日期：2018-12-11

作者简介：王兴宇，厦门大学高等教育质量建设协同创新中心/厦门大学教育研究院博士生。（福建厦门/361005）

* 本文系教育部哲学社会科学研究重大课题攻关项目“高校人才培养质量保障研究”（16JZD045）的成果之一。

决策支持型信息,这也说明美国高等教育管理信息系统已经处在智能化决策支持的过程中。另外,IPEDS 会按照不同利益主体的需求和认知水平将数据进行相应的分类处理并在网站上进行公布。联邦政府、州政府、学校、教育研究者、媒体乃至第三方教育机构等都可以根据自身诉求,使用 IPEDS 提供的数据工具获取全国、全州或某一类高校的数据资料和统计结果。^[4] 这为教育发展中热点和难点问题的把握提供了科学的依据,也为高等教育管理提供了参考和决策的支持,提高了全美高校教育信息管理的实时反应能力。由此可见,IPEDS 已经发展成为政府、学校、学者、公众等利益相关者描绘高等教育轮廓的“画板”,成为全美集教育信息管理、科学研究、教育决策于一体的高等教育数据综合体和信息中心。

二、美国高等教育综合数据系统的运行

IPEDS 是一个多维、动态的数据系统。国家教育统计中心 (National Center for Education Statistics, NCES) 从数据定义、收集、使用、公开、储存等多个方面对其进行了科学设计,使其在数据统计的规

范性、数据来源的整体性、数据使用的时效性等方面得到充分保障,零碎的、分散的高等教育数据资料得以整合成规模化、结构化的数据系统,使得 IPEDS 成为美国高等教育数据管理的门户。而对于我国高等教育数据系统的建设和发展来说,比较关注的是 IPEDS 的数据来源、采集和使用等问题。

(一) 数据来源

在数字时代,数据是教育发展的基础设施,蕴藏着巨大的潜能,是高等教育数据系统的核心资源。经过长期发展和不断修正,IPEDS 所收集的数据基本涵盖高等教育的所有方面,涉及高等学校的机构特征、入学、学位授予、毕业率和学习成果测量、招生、学生资助、人力资源、财政、图书馆等各个领域。^[5] 参与调查的学校基本包括所有不同类型的学院和大学,遍布美国 50 个州和其他司法管辖区^①。2016 - 2017 年有 6760 所高校完成向 IPEDS 的数据提交工作 (如表 1 所示)。这些学校不仅包括 4 年制的研究型大学、州立大学,还包括 2 年制的文理学院、社区学院等高校,同时也包括一些学制不足 2 年的技术学院、美容学院等非学位授予的高校。

表 1 参与 IPEDS 数据调查的高校数量和百分比分布汇总表 (2016 - 2017 年)

学校类型和所在地区		高等教育机构数量				高等教育机构百分比 (%)			
		总计	公立	私立		总计	公立	私立	
				非营利	营利			非营利	营利
学校类型	总计	6760	1985	1876	2899	100.0	100.0	100.0	100.0
	4 年制美国本土高校	2839	737	1588	514	42.0	37.1	84.6	17.7
	4 年制其它美国司法辖区高校	79	18	46	15	1.2	0.9	2.5	0.5
	2 年制美国本土高校	1969	981	158	830	29.1	49.4	8.4	28.6
	2 年制其它美国司法辖区高校	26	9	5	12	0.4	0.5	0.3	0.4
	少于 2 年制美国本土高校	1798	240	77	1481	26.6	12.1	4.1	51.1
	少于 2 年制其它美国司法辖区高校	49	0	2	47	0.7	0.0	0.1	1.6
所在地区	新英格兰地区	388	109	157	122	5.7	5.5	8.4	4.2
	中东部地区	1082	275	410	397	16.0	13.9	21.9	13.7
	五大湖地区	968	264	287	417	14.3	13.3	15.3	14.4
	普莱恩斯	570	187	184	199	8.4	9.4	9.8	6.9
	东南部地区	1665	541	394	730	24.6	27.3	21.0	25.2
	西南部地区	721	241	113	367	10.7	12.1	6.0	12.7
	落基山地区	266	81	43	142	3.9	4.1	2.3	4.9
	远西区	941	255	235	451	13.9	12.8	12.5	15.6
	军事院校	5	5	0	0	0.1	0.3	0.0	0.0
	其它司法辖区	154	27	53	74	2.3	1.4	2.8	2.6

数据来源: NCES. 2016 - 17 Integrated Postsecondary Education Data System (IPEDS) Methodology Report [EB/OL]. [2018 - 04 - 06]. <https://nces.ed.gov/pubs2017/2017078.pdf>.

目前, IPEDS 具体的数据收集工作主要从 11 个调查部分展开。(1) 院校特征。院校特征数据主要收集每所院校的名称、地理位置、教育服务、提供远程教育的机会、控制或隶属关系、入学要求以及学生服务等基本信息。另外, 学生的消费, 如学费、书费、食宿和其他费用(如交通费、洗衣费和娱乐等)数据也会在这一部分的调查中被收集。学院和财政援助办公室会根据学生的消费数据来确定其是否需要资助。(2) 学位授予和证书获得。这一调查部分主要收集院校授予的学位、学生获得的证书和奖项的数量。IPEDS 会根据奖项的级别、种族/民族、性别以及学科专业目录(CIP)对获得相应证书或学位的学生群体进行分类统计。(3) 年度招生数。因为美国的高校之间可以进行学分互认, 而且允许学生自由转学, 所以年度招生数指的是 12 个月内不重复进入一所高校的学生数量。为了保证数据的准确, IPEDS 同时会收集教学活动数据(如签到等), 并据此计算每年全日制本科生和研究生的入学人数。(4) 学生资助。这一部分主要收集每所院校受资助的学生数和相关资助的额度, 具体关注点主要集中在全日制学生和大一新生以及其所受资助的类型和学生的平均资助率。IPEDS 会将其作为计算学生平均花费的重要参考, 并确定院校和学生应获得的拨款和资助金额。(5) 毕业率。在美国毕业率是一个非常重要的高等教育概念, 是院校分配教育资源、学生选择学校、媒体评估教育环境的重要指标。在这一部分中, IPEDS 主要收集院校全日制学生数、初次申请学位的学生数、转学生数以及 100% 毕业率^②、150% 毕业率和 200% 毕业率。(6) 成果测量。学生学习成果测量是 IPEDS 在 2016 年最新加入的一个调查部分, 它要求各院校提供四组学生的学习成果数据, 即全日制在校生、全日制毕业生、非全日制在读生和非全日制毕业生。(7) 入学申请。入学申请是院校招生部门选拔本科生考虑因素的具体体现。在这一部分, IPEDS 主要收集参与申请的学生数、学生的入学考试分数以及中学期间的档案袋记录。(8) 秋季注册。这一调查部分主要是收集新入学学生的人口学数据, 主要包括本科生和研究生总人数、性别、年龄、种族以及其是否需要远程学习、是否为全日制、是否寄宿等信息资料。另外, IPEDS 也会要求院校提交新一年的学生留存率和师生比的数据。(9) 财务概况。财务数据主要涉及院校的年度总收入和总支出、净资产变化、捐

赠以及奖学金和研究经费。但是由于高校类型(公立、私立营利、私立非营利等)的不同, 其财务执行的标准也不尽相同。通常情况下, 财务数据的提交会使用政府会计准则委员会(Governmental Accounting Standards Board, GASB)或者财务会计准则委员会(Financial Accounting Standards Board, FASB)的标准。(10) 人力资源。高等院校的教员工体系的差异性很大, IPEDS 主要从人口学和职业特征来收集全职教学研究人员、全职非教学研究人员、兼职人员和新入职的全职人员的基本信息, 其中包括工作类别、职位、学历、性别、种族等数据资料。(11) 学术图书馆。IPEDS 对图书馆数据的收集主要由两部分组成: 其一为图书馆馆藏、馆际互借服务、数据库和电子媒体的使用量、图书馆借阅量(包括纸质书、媒体、电子书)、阅览室上座率等核心数据; 其二为图书馆员工薪资和福利、图书购买、馆内日常运营和维护支出等管理数据。经过多年的积累和更新, IPEDS 已经形成了涉及不同教育主题、不同教育领域的横向纵向数据, 数据完整性高、覆盖范围广、时间跨度长的教育综合数据系统, 逐渐成为支持高等教育政策制定, 提供及时、准确的数据信息的“教育利器”。

(二) 数据采集

IPEDS 是年度数据收集项目, 每年都会对数据进行更新。由于不同高等教育数据产生的时间不同, 例如招生数据需要在学生入学后才能获得, 图书馆数据则需要在校年结束后进行汇总, 为了保证数据能够及时快速地得到更新, IPEDS 每年分秋季、冬季和春季 3 个时期收集高等教育数据。其中, 院校特征、学位授予和证书获得、年度招生数为秋季数据, 需要各参与高校在每年的 9 月份完成提交; 学生资助、毕业率、学习成果汇报、入学申请信息为冬季数据, 完成提交时间为 3 月份; 春季数据为财务、人力资源、秋季注册信息、学术图书馆数据等, 各高校需在 4 月份将其提交给 IPEDS。^[6]

IPEDS 的数据来源高校数量多、涉及范围广, 所以数据收集的难度非常大。为了能够保证数据收集工作的有效开展, IPEDS 专门设计了数据管理员、数据协调员、数据中心三级高等教育数据协调收集制度(如图 1 所示)。IPEDS 要求所有参与数据提交的高校必须推荐一名数据管理员, 负责确保其所在高校提交的数据的完整性和可靠性。在数据收集工作中, 数据管理员需要对高校的数据提交工作进行

指导; 数据提交完成后, 管理员需要利用 IPEDS 提供的密钥进入数据系统对数据进行核查、锁定和上传。IPEDS 按照地区、学校类别等因素对所有参与数据提交的高校进行了分组, 每组高校至少有一名数据协调员负责对数据管理员提交的数据进行再次审查, 并针对数据中存在的问题与管理员进行沟通。一旦数据被协调员锁定和上传, IPEDS 的数据中心会对其进行最终审查。如果数据存在问题, 数据管理员和协调员会收到相应的数据修订通知, 而最终通过三级审查的数据会被迁移到 IPEDS 的数据中心。

同时, 为了保证数据统计的科学性、结构化和后续处理的适用性, IPEDS 为参与调查的高校提供了在线数据录入表单。这是一套根据高校性质 (公立、非营利、营利)、高校类型 (4 年制、2 年制和 2 年制以下)、是否授予学位以及一些学术条款等因素为高校量身定制的数据录入系统 (如图 1 所示)。通常情况下, 在线数据录入表单系统会预先加载前几年的数据, 以便于高校参考和比较。例如学位授予的数据需要按照系统预先加载的 CIP 代码进行提交, 其数值的增减幅度一般不会超过 50%; 对于学生费用数据而言, 增减幅度要更小, 食宿费用为 40%, 学费为 30%。^[7] 如果因为特殊情况, 数据波动较大, 高校可以在表单录入系统提供的背景框中进行解释和说明, 否则数据将不能通过系统的审核。

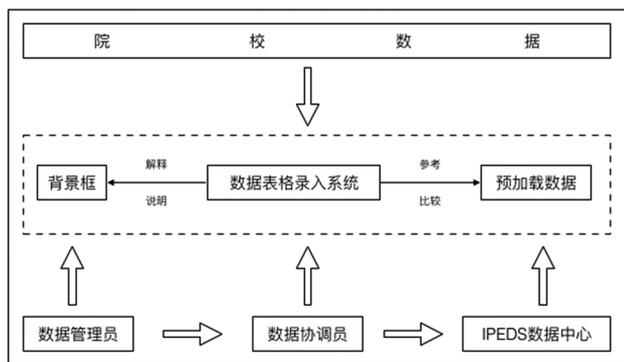


图 1 IPEDS 数据收集和审查结构图

(三) 数据利用

数据的价值在于应用, 教育数据的价值在于满足不同利益相关者了解和影响教育的需要。IPEDS 通过在其数据中心主页进行公布和发布各种教育报告等途径向公众公开, 以满足各利益相关群体的不同教育诉求。其具体数据发布流程主要分为初步数据、最终数据、最终数据的修订数据三个阶段。通常在数据收集周期结束后, IPEDS 便会发布初步数

据和基于该数据的“First Look”教育报告。^[8] 一方面是为了保证数据的时效性, 另一方面是为了收集不同利益相关者对数据的诉求以及对数据中存在问题的反馈。在经过第一阶段的反馈和国家教育统计中心 (NCES) 的所有数据质量控制程序后, IPEDS 将发布最终数据和“First Look”教育报告的正式版。在最终数据发布之后, IPEDS 数据中心仍会接受用户的反馈和高校对数据的修订, 但是这些信息通常只会被用来作为后续数据收集的指导和参照。

IPEDS 在公开发布数据的同时也在其主页上提供了多种针对不同群体的数据使用工具。例如, 帮助学生和家长了解大学整体概况, 对不同高校进行查找和比较的学院导航 (College Navigator), 以研究者为中心开发的数据访问工具“USE THE DATA”以及快速获取某一所高校综合资料的发现大学工具 (Look Up an Institution) 等。^[9] 同时, IPEDS 也提供了国家、地区、州以及高校等多个层面的数据趋势和统计表, 以帮助联邦政府、州政府和高校更快速、更科学地获得所需要的教育管理和决策证据。

另外, 为了深入挖掘数据的价值, 国家教育数据统计中心通过资助院校研究协会 (Association For-Institutional Research, AIR) 对 IPEDS 数据的获取和使用进行培训, 使得 IPEDS 为更多人所关注、熟知和使用。AIR 针对 IPEDS 的每个数据组成部分分别开发了不同的数据获取和使用的视频教程, 并且随着 IPEDS 数据的更新进行定期的调整。^[10] 除此之外, 每年 AIR 都会举办至少 30 场工作坊, 即面对面的数据使用培训课程, 其主题包括 IPEDS 数据调查的各个组成部分。例如“IPEDS 人力资源培训”“IPEDS 数据: 高校的公众形象”“IPEDS 数据: 教育决策和管理效能的保障”等。工作坊通常由数据提供者和用户共同参与, 研讨式的培训设计除了提供对 IPEDS 数据工具使用和演示的基础训练, 还旨在促进各方利益相关者对数据定义的理解, 提升其数据管理技术。^[11]

三、美国高等教育综合数据系统的特点

经过多年的发展, IPEDS 在美国高等教育信息化的过程中发挥了重要作用, 已经是一个能够承载多元功能的高等教育综合数据系统。通过对整个系统的分析发现, IPEDS 在数据的完整性、可获得性、利用价值等方面特点显著。

(一) 整体规划, 保证数据的全面完整

教育数据是对教育系统及其活动和活动结果的记录。由于现代高校规模日渐庞大、功能更加多样、结构愈益复杂, 其产生和积累的数据也愈加繁杂和混乱, 加之缺乏统一的数据规划和数据标准, 数据之间流通不畅, 使得“数据孤岛”成为教育领域的普遍现象。IPEDS 利用高等教育修正案的强制性力量确定数据的来源, 通过顶层设计对数据进行统一规划和统一收集, 从而使数据的全面和完整得到了保障。此外, 已有研究也表明, IPEDS 的丰富数据服务于各方面的人群, 满足多方面的利益相关群体的需要, 如满足院校对法律文本的强制性要求、服务求学家庭、服务院校相关研究人员等。^[12]这说明 IPEDS 数据的全面性不仅体现在数据来源和构成方面, 还体现在数据的深度应用方面。1992 年美国高等教育修正案在学生资助计划参与协议中增加了关于美国高等教育综合数据系统 (IPEDS) 的规定, 要求所有参与联邦学生资助计划的高校必须协助完成高等教育综合数据系统 (IPEDS) 的数据收集工作。^[13]参与该计划的高校达 7000 所之多, 几乎涵盖了美国所有不同类型、不同层次、不同地区的高校, 这使得数据的来源范围空前扩大, 数据量空前增加, 不同高校的有价值的的数据得以整合, 从来源上保证了数据来源的全面和完整。在数据构成方面, IPEDS 对高等教育内部数据进行总体构思, 统一规划, 梳理出 11 个数据组成部分, 既包括从入学到毕业的整个教育过程的数据, 也包括机构特征、学术图书馆等教学辅助数据, 使得数据的综合性和整体性得以提高。在数据应用方面, IPEDS 考虑到所有利益相关群体的知识结构、认知水平的不同, 对数据诉求的不同, 开发了不同的数据分析工具, 并对数据进行了不同的处理, 以满足多样化的需要, 提升数据的利用价值。

数据是对事物和现象的客观描述, 高等教育数据是对高等教育现象的客观描述。在某种程度上, 教育数据就是教育系统和教育活动的“镜像”, 如果数据的完整性和全面性缺失, 透过数据对教育现象的描述、教育问题的发现、教育决策的制定则无法保证其科学性。IPEDS 在全局视野下对教育数据进行通盘考虑: 分析数据需求信息、确定数据目标体系、设计数据结构, 整体规划, 使之成为统一、集中、稳定的数据集。这不仅保证了教育数据的整体性, 同时也为后续数据采集工作打下了良好

的基础。

(二) 三级管理, 确保数据的准确客观

客观准确是数据最重要的属性之一。数据失实, 其使用价值也会随之丧失。其准确性受到对数据的理解、数据的收集、数据的存储等多重因素的影响。因此, 要想保证教育数据的客观准确, 不仅要有科学合理的数据系统架构设计, 还要有全方位的监督和审查。为了避免高校陷入收集数据、整理数据、提交数据的疲劳战, 保证数据采集有计划、有分工、有步骤地开展, IPEDS 考虑到不同教育数据产生时间和周期的不同, 每年分三个时期收集高等教育数据。这一方面保证了 IPEDS 数据及时快速更新, 另一方面新产生数据的直接提交可以避免在个人或组织处停留而掺杂过多不确定因素而降低数据的质量。

为了避免对数据产生误解和收集过程出现影响准确性的因素, IPEDS 开放并建立了数据录入表单系统以及数据管理员、协调员和数据中心三级数据收集管理制度。数据表单是根据不同类型、不同层次高校的具体数据情况定制的数据录入系统, 这一做法的目的是在数据采集前期通过规则设计预防低质量数据, 通过该表单系统帮助高校提升所提交数据的结构化程度, 优化数据的汇总流程。数据提交后, 首先由每所学校的数据管理员和负责州或地区部分高校的数据协调员进行审查和锁定, 协调员通常是在数据管理员的基础上进行高密度的抽查。^[14]也就是说, 数据在到达 IPEDS 数据中心之前要先进行两次“安检”。IPEDS 数据中心会对数据进行最终的审查。整个的数据提交流程中, 从院校到数据中心, 各级负责人员会不断进行沟通, 并最终在对数据的理解上达成一致。不难发现, IPEDS 在数据采集前的规则设计和收集中的多级审查可以有效地控制影响数据质量的不利因素, 提升数据质量, 是一套比较严密的数据采集管理规范。

(三) 赋权增能, 保障数据的利用价值

数据的价值体现在数据应用。教育数据本身并没有意义, 只有与教育情景紧密联系和整合才能成为信息, 并最终转化为知识。^[15]例如, 雇主根据学校综合数据分辨各类文凭的价值, 为业务的拓展选择更优质的劳动力; 管理者根据教育运行数据, 支持教育决策; 教师根据学习成果数据, 为学生制定教学方案, 提升学生学习体验。为此, IPEDS 非常重视面向利益相关者的赋权增能, 即最大限度地赋予各利益群体访问、获取和使用数据的权利; 通过

数据分析工具开发和系统培训提高利益群体对数据的理解和使用数据的能力。

对不同利益相关者的支持要建立在对其数据的理解和描述的基础上。^[16] IPEDS 考虑到不同利益相关者拥有不同的数据需求和不同的数据使用水平,开发了各种基于角色的数据服务工具。例如数据趋势生成器、分变量汇总表、可视化数据报告等可以根据不同利益相关者的需要生成相应的学院和学生表现数据,从而帮助管理者优化资源配置,帮助学生和家庭做出理性的教育选择,帮助教师和行政人员改进教学服务。另外,这些数据工具也是在数据开放共享和隐私保护之间做出的平衡。根据美国的数据隐私保护法,除学生、家长和管理员外,其他人无权访问带有学生身份识别信息的数据,而对于不包含身份识别信息的汇总数据,所有利益相关者均有权获得。^[17] 为了更有效地发挥 IPEDS 数据的价值,国家教育统计中心非常重视教育数据研究者和教育工作者数据素养的提升,并资助院校研究协会开展了一系列针对教师、管理者、雇主、董事会等群体使用数据进行教育决策所需的知识、技能和行为的培训,力图使教育数据能够与教育情景更为紧密的结合起来,发挥出数据的潜在价值。

四、美国高等教育综合数据系统的启示

在信息时代,数据已经成为推动教育变革的新型战略资产和科学力量。^[18] 我国政府高度重视教育数据资源建设,2014 年发布《教育管理信息化建设与应用指南》,将数据的深度挖掘和应用作为未来工作的重点;2015 年开始战略部署教育大数据,成立中国教育大数据研究院、教育大数据应用技术国家工程实验室等研究机构,将教育数据作为“十三五”期间重点发展的内容。^[19] 在教育管理信息化、数据化的教育新时代,我国可以从顶层设计、技术引进、分析应用和文化培育等方面借鉴美国高等教育数据系统建设的经验。

(一) 顶层设计: 做好教育数据系统的总体规划

由于缺乏统一规划和主数据管理系统,高校内部的数据通常都是分散维护。例如招生和就业数据由招生办和就业指导中心分别掌握,财务和评估数据也隶属于不同的部门。数据归属的分布性和数据管理的碎片化给教育数据系统的建立和数据的收集带来了巨大困难。IPEDS 正是意识到教育数据的分散性,以跨部门、跨功能、跨系统的数据流为切入

点,从战略的高度制定了顶层框架,对数据内容构成、数据收集流程、数据挖掘与使用做出整体性规划,以保障整个数据系统的有序健康运行。

事实上,高校中存在的数据横向流通受阻、纵向传递不畅、相关关系混乱等问题是建设教育数据系统必须要攻克的难题。借鉴 IPEDS 的经验,我国建设教育系统,提升高等教育管理的科学化程度,可以从规划数据内容、培育数据管理人员、设立数据咨询机构等方面进行顶层设计。通过数据内容的整体规划为分散的数据建立联系,提升数据的整体性和全面性,为后续的收集提交做好铺垫;通过整体培育和布局数据管理人员,保证各个院校、各个地区的数据收集工作有序推进;在此基础上发挥教育咨询机构的作用,促进不同利益相关者对数据的理解,保证数据的准确性和内部一致性。最终建立起多级数据管理人员、数据咨询机构、信息化部门、院校师生等多个数据利益群体认同的协同数据环境。

(二) 技术引进: 保障教育数据系统的有效运行

从源头上看,数据存在的问题——归属的分布性,管理的碎片化等——在很大程度上是技术的问题。为此,IPEDS 已和多所高校合作引进了以数据仓库、数据挖掘和联机分析等为核心的商业智能技术(Business Intelligence, BI),帮助高校控制数据质量,聚合业务数据,保障数据内部流通和联系的通畅。我国高等教育规模位居世界第一,每天的数据产出量盈千累万,这对数据的管理提出了极高的要求。高校作为大规模数据的产出方,虽然掌握着教育数据的使用权,但是数据管理技术相对落后,唯有引进和创新教育数据管理技术,方可拥有主动权。我国高校可以参考 IPEDS 的经验,与商业领域合作,引进外部数据管理技术,例如商业智能(BI),以保证高校内部数据的流通和使用。目前,我国的商业智能技术已经走在了世界的前列,例如阿里巴巴、腾讯等企业已经在 BI 领域做出了杰出的成就。通过将 BI 技术引入教育领域,在高校内部建立主数据中心,形成数据仓库,以业务数据系统的数据为基础,整合数据并将其转化为有序的信息,可以有效避免数据孤岛、数据冗余、数据标准与内容不一致等传统数据管理方法存在的缺陷,从而将分散、零碎的数据结构化为一个整体,形成准确、完整、一致的数据库和便捷、合规的分享、访问、应用高等教育数据的技术环境。

数据技术引进的目的是在整体数据管理水平偏低的情况下,对教育数据进行合理、有效地重组,最终满足高校的实际需求,从而产生数据分析和挖掘的价值。所以说,技术是保障数据价值得以发挥的关键因素。显然,BI是非常优秀的管理技术,但是对于实现高校数据的科学管理、保障高校教育数据系统的有效运行而言,BI并不是唯一的选择。在2012年10月美国教育部发布《通过教育数据挖掘和学习分析促进教与学》的报告后,数据挖掘和学习分析技术逐渐发展成为学界公认的教育大数据处理技术。尽管基于数据的教学行为优化和学习行为监测并不是新鲜事物,但是教育数据挖掘和学习分析技术的出现,可以实现对于大量细粒度教育数据的采集、处理和分析,从而为教育机构的教育决策、教师的教学决策、学生的自我学习监控提供更加精细化的服务。^[20]重视和引进数据挖掘和学习分析技术势必可以提升数据管理的科学性和保障系统运行的有效性。

(三) 分析应用: 推进高等教育治理的科学决策

教育数据系统的建立归根结底是要促进教育管理的信息化、科学化和数据化。美国的实践经验表明,数据的最终价值体现在与教育情景的紧密联系和充分整合,这也是实现高等教育的理性决策和科学管理的必由之路。相关研究也表明,加强对数据的统计分析与应用,向公众发布高等教育状况统计报告已经显示出一定的时代紧迫性。^[21]现代大学已经处在社会中心,不再是“不问世事”的“象牙之塔”,来自政治、经济、文化等因素的影响使大学组织愈加复杂,也对其管理提出了巨大挑战。时代已经证明,以数据分析和数据应用为基础,调和高等教育中的冲突和矛盾,弥合教育业务领域的空白,消除管理的盲点,是现代大学治理的有效选择。

随着我国高等教育改革的逐步深入,大学排名、学科评估、绩效考核、社会问责等现实问题在高等教育中愈发突出。我国高校应借鉴美国经验,遵循信息时代的规律,尊重数据,树立正确的“数据观”,发展以数据为基础的思维方式,建立数据开放共享机制,推进数据化战略,将相关资源进行数据化处理,盘活已有数据存量,利用数据增量,进行数据挖掘和应用,进而提高高等教育的管理效率,增强高等教育决策的科学性,提高教育质量,促进教育健康发展。^[22]

(四) 文化培育: 营造重视教育数据的良好氛围

数据不仅是工具和资源,还是一种重要的战略和文化。“数据文化”的根本在于价值理念和思想观念。数据是在事实或观察的基础上对客观事物作出的逻辑归纳。它是一种尊重事实、强调精确、推崇理性的文化。^[23]IPEDS之所以能够成为美国高等教育的数据中心,成为监测教育质量、科学决策的“利器”,其主要原因在于美国长期以来强调实证精神、重视数据所形成的深厚数据文化土壤为其创造了良好的发展环境。这在IPEDS数据收集、分析和应用的整个过程中也有非常明显的体现。例如院校对数据收集的支持,公众对数据公开和使用的要求,研究者对数据获取的建议等,都说明数据文化已经成为美国高等教育信息管理机制建设的重要保障。

我国高校经过十几年的信息化建设,虽然已经积累了大量数据,但由于数据文化的缺失,教育数据的开发和利用程度还处于初级阶段。这需要我们加大“数据文化”培育的力度,让关注数据、分析数据、使用数据成为学生、教师、教育研究者等群体的内在习惯。文化的形成是一个长期的过程,数据文化的培育亦是如此,需要久久为功的数据意识培养和数据制度建设。数据意识的培养有赖于各级教育部门和高校加强对教育数据的宣传和引导,并鼓励各高校举办相关论坛,奖励相关科研成果,从而推进教育治理现代化、信息化、数据化。而数据制度的建设则是要对数据的分享、访问和使用进行制度规定,建立数据的共享机制、隐私保护制度等,为教育研究者、公众、高校等利益相关者分享和使用教育数据保驾护航。通过数据意识培养和数据制度建设,提升数据文化的内涵,在高等教育领域的海量数据中建立数据体系、标准和系统,真正发挥出数据文化的作用。

注释:

①IPEDS将高等院校定义为向公众开放,并提供高中以上教育和培训的组织,包括提供学术的、职业的和继续教育项目的高等教育机构。

②众所周知,美国高校一直是“宽进严出”,所以其毕业率往往不高。学生无法在正常年限毕业也非常正常。美国大学对毕业率的计算往往分为4年毕业率,6年毕业率等。4年毕业率也叫100%毕业率,也就是学生花费了学制的时间100%取得毕

业。同理,6年毕业率,就是学生花费了正常学制150%的时间才取得毕业,所以也叫150%毕业率。

参考文献:

[1]NCES. 2016 - 17 Integrated Postsecondary Education Data System (IPEDS) Methodology Report [EB/OL]. [2018 - 04 - 03]. <https://nces.ed.gov/pubs2017/2017078.pdf>.

[2]NCES. IPEDS Data Feedback Reports [EB/OL]. [2018 - 04 - 03]. <http://discover.alpenacc.edu/reports/ipeds.php>.

[3]NCES. IPEDS Media Primer [EB/OL]. [2018 - 04 - 03]. <https://nces.ed.gov/pubs2017/2017047.pdf>.

[4]IPEDS Data Center. Use the Data [EB/OL]. [2018 - 04 - 05]. <https://nces.ed.gov/ipeds/use-the-data>.

[5]NCES. Getting Started with the IPEDS Data [EB/OL]. [2018 - 04 - 06]. https://nces.ed.gov/training/datauser/IPEDS_02.html.

[6]NCES. Data Collected Through the IPEDS [EB/OL]. [2018 - 04 - 06]. https://nces.ed.gov/training/datauser/IPEDS_03.html.

[7]NCES. 2016 - 17 Integrated Postsecondary Education Data System (IPEDS) Methodology Report [EB/OL]. [2018 - 04 - 08]. <https://nces.ed.gov/pubs2017/2017078.pdf>.

[8]NCES. Overview of IPEDS Data Release [EB/OL]. [2018 - 04 - 08]. https://nces.ed.gov/ipeds/resource/download/IPEDS_DataReleaseProcedures.pdf.

[9]NCES. IPEDS Data Tools [EB/OL]. [2018 - 04 - 07]. <https://nces.ed.gov/datatools/>.

[10]AIR. IPEDS Educators [EB/OL]. [2018 - 04 - 07]. <http://www.airweb.org/educationandevents/ipedstraining/pages/ipedstrainers.aspx>.

[11]AIR. Face - to - face Workshop [EB/OL]. [2018 - 04 - 07]. <http://www.airweb.org/Education-AndEvents/IPEDSTraining/Workshops/Pages/default.aspx>.

[12]屈琼斐,舒忠梅.美国中学后教育数据综合系统IPEDS的管理与运行机制研究和启示[J].高教探索,2013(5):82-87.

[13]NCES. The History and Origins of Survey Items for the Integrated Postsecondary Education Data System (2016 - 17 Update) [EB/OL]. [2018 - 04 - 07]. <https://nces.ed.gov/surveys/SurveyGroups.asp?Group=2>.

[14]NCES. IPEDS Statutory Requirements, Data Collection Procedures, Universe, Levels of Analysis, and Missing Data [EB/OL]. [2018 - 04 - 08]. https://nces.ed.gov/training/datauser/IPEDS_04.html.

[15]彭雪涛.美国高校数据治理及其借鉴[J].电化教育研究,2017,38(6):76-81.

[16]胡钦太.高等教育信息化深度发展框架与趋势分析[J].教育研究,2009,30(10):97-101.

[17]Data Quality Campaign. Student Data Privacy Legislation: A Summary of 2016 State Legislation [EB/OL]. [2018 - 04 - 08]. <https://dataqualitycampaign.org/resource/2016-student-data-privacy-legislation/>.

[18]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.

[19]李振,周东岱,刘娜,董晓晓,钟绍春.教育大数据的平台构建与关键实现技术[J].现代教育技术,2018,28(1):100-106.

[20]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,2013,31(6):11-17.

[21]张艳丽.美国国家教育统计中心与高等教育数据运行机制的探索[J].高等理科教育,2016(1):80-85+106.

[22]许晓东,王锦华,卞良.高等教育的数据治理研究[J].高等工程教育研究,2015(5):25-30.

[23]涂子沛.数据之巅:大数据革命,历史、现实与未来[M].北京:中信出版社,2014:1.

(责任编辑 陈志萍)