

# 美国高等教育数据库的建设及其启示

王 竞, 胡咏梅

**[摘要]** 美国联邦教育部非常重视教育数据资料的采集工作, 采集数据资料具有悠久的历史。本文阐释了美国联邦教育部建设的高等教育数据库(IPEDS)的结构和功能, 以及其数据采集制度。此外, 概述了AAUP、NSSE等几项高等教育专题数据库的结构和运作方式, 最后, 给出了对我国高等教育数据库建设的若干启示: 一、依据所需建设的数据库的性质区分数据库建设主体; 二、依据数据库建设的目的确立各类高等教育数据库的功能; 三、系统规划各类高等教育数据库的结构和指标体系; 四、建立数据库共享机制, 最大程度的发挥各类高等教育数据库资源的社会效益

**[关键词]** 美国高等教育数据库(IPEDS); 数据库结构; 数据库功能; 数据库建设

美国联邦教育部采集资料具有悠久的历史。早在1867年建立时, 它的主要使命就是“采集统计数据 and 资料, 以了解现有各州学校的现状和进展, 向公众介绍学校的组织管理, 学校制度和教学方法, 使公众明了建立和保持一个高效率的学校系统的重要性, 以促进整个国家的教育发展”(Aliyeva, et al., 2018)。尽管联邦教育部曾一度被并入其他政府部门, 但资料采集工作并没有停止。

联邦教育部的资料采集是一个渐进的过程: 1869—1870年开始采集学生入学、学位授予和教学人员的数据, 1870—1874年增加了图书馆和财务的数据。1974年, 全美教育统计中心(National Center for Education Statistics, 以下简称NCES)正式成立, 是隶属于联邦教育部的一个部门, 负责数据和资料的采集工作。这一年, 增加了学费、教员工资、非教学人员、学生资助和研究生等项目。而大规模系统化的高等教育资料采集, 即高等教育数据库

**[收稿日期]** 2019-09-20

**[基金项目]** 国家自然科学基金面上项目“‘双一流’建设背景下高校教师工资制度改革研究”(71874015)、北京市社科基金研究基地重点项目(18JDJYA003)。

**[作者简介]** 王竞, 美国加州州立大学萨克里门托分校行政管理研究处, 电子邮箱地址: jwang@csus.edu; 胡咏梅, 北京师范大学教育学部/首都教育经济研究院, 电子邮箱地址: huym0718@bnu.edu.cn。

(Integrated Post-secondary Education Data System, 以下简称 IPEDS)则是从1986年开始建设的。先是分类采集数据和资料,自1990年起合为一个统一的采集过程。此后,又进行了多次的修改补充。1998年,IPEDS改用互联网采集数据,提高了采集的速度和准确性,并增加了多种检索功能。

为了保证资料采集的准确性,联邦教育部发布了一系列的规则法令,例如,规定全国统一的学科名称(Classification of Instructional Programs, 简称 CIP)<sup>①</sup>以及确立种族划分的主要规则。其中最重要的一条法令,是于1992年将 IPEDS 数据采集列为高校学生资助法案的一个补充条款,使 IPEDS 数据采集具有了法律效力<sup>②</sup>。

本文结构安排如下:第一部分介绍美国高等教育数据库的结构;第二部分阐述该数据库的五项主要功能;第三部分说明该数据库的建设条件和数据采集制度;第四部分概述几个重要的高等教育专题数据库的结构信息、规模范围和运作方式;第五部分阐释美国高等教育数据库建设对我国高等教育数据库建设的几点启示。

## 一、美国高等教育数据库的结构

本文以2018版的IPEDS为基础,对美国高等教育数据库进行介绍和分析。下面是IPEDS资料采集的12个项目一览表<sup>③</sup>:

表1 美国联邦教育部高等教育数据库的结构

项目	指标	算法及要求	备注
1. 在校生	<ul style="list-style-type: none"> <li>• 全年在校生总数</li> <li>• 学生特征: 本科生和研究生, 性别和种族</li> <li>• 学分总数</li> <li>• 全日制学生当量总数(FTE)</li> </ul>	年度全日制学生当量总数(FTE): 用学分计算: 本科生=30学分; 研究生=24学分 用人数计算: 半日制本科生=0.34-0.40; 半日制研究生=0.36-0.38	FTE: Full Time Equivalent

① <https://www.carnegieclassifications.iu.edu>。

② 引自美国法律: 20 USC 1094, Section 487(a)(17) and 34 CFR 668.14(b)(19), <https://nces.ed.gov>。

③ <https://nces.ed.gov>。

续表

项目	指标	算法及要求	备注
2. 图书馆	<ul style="list-style-type: none"> <li>• 图书, 报刊杂志, 丛书和检索数据库的总数以及电子版总数</li> <li>• 图书和报刊杂志的流通量和电子媒介的使用率</li> <li>• 馆际图书和报刊杂志的流通量</li> <li>• 图书馆年度支出</li> </ul>	图书馆年度支出低于 10 万美元不必上报	
3. 招生标准和数据	<ul style="list-style-type: none"> <li>• 招生标准细则</li> <li>• 接收率</li> <li>• 入学率</li> <li>• SAT 或 ACT 平均分数</li> </ul>	接收率 = 接受人数 / 申请人数 入学率 = 入学人数 / 接受人数	
4. 学位授予	<ul style="list-style-type: none"> <li>• 各学科专科, 本科和研究生学位授予数, 分性别和种族</li> <li>• 获学位的学生人数, 分专科, 本科和研究生学位, 分性别和种族</li> <li>• 含联网学位人数</li> </ul>	根据美国联邦教育部颁发的统一的学科名称和编号(CIP), 填写本校的各个学科的数据	CIP: Classification of Instructional Programs
5. 秋季新生数	<ul style="list-style-type: none"> <li>• 全日制和半日制新生人数(专科, 本科和研究生学位, 性别和种族)</li> <li>• 新生的籍贯和是否持有高中文凭</li> <li>• 新生的平均年龄</li> <li>• 新生第一年保持率</li> <li>• 教师/新生比例</li> <li>• 含联网学习人数</li> </ul>	同年秋季入学的全日制本科新生定义为同级组(Cohort)。 Cohort: 同年入学的学生 Class: 同届毕业的学生	
6. 学校财政	<ul style="list-style-type: none"> <li>• 年度收入</li> <li>• 年度支出</li> <li>• 不动产和保险金</li> <li>• 研究资金和奖学金</li> </ul>	年度收入, 包括学费, 杂费, 政府提供的研究资金和合同, 私人捐款总额 年度支出, 分为教学, 科研, 教辅和服务等	

续表

项目	指标	算法及要求	备注
7.4—6年 毕业率*	<ul style="list-style-type: none"> <li>• 同级组人数</li> <li>• 性别和种族</li> <li>• 是否获低收入助学金 (Pell Grant)</li> </ul>	毕业率 = 4年或6年毕业人数/同级组人数	
8.8年 毕业率	• 与上一项相同,但使用不同的同级组	毕业率 = 8年毕业人数/同级组人数	新加的项目
9. 教职工	<ul style="list-style-type: none"> <li>• 全日制和半日制教职工总数</li> <li>• 全日制教辅人员的级别, 性别, 种族, 合同期</li> <li>• 全日制教学人员的级别, 性别, 种族, 合同期, 工资总额</li> </ul>	教职工总数, 要标明多少人具有教学人员身份	美国大学有些行政人员兼课或保留教授身份
10. 学校基本信息	联系人, 学生服务机构, 学费杂费及其它费用, 公立或私立, 2学期或4学期制, 学位授予权和提供的学科等		
11. 其它教育成果	<ul style="list-style-type: none"> <li>• 半日制新生和转校生8年内学位获取人数</li> <li>• 半日制新生和转校生第九年的入学状况</li> </ul>		
12. 学生的奖学金和助学金	<ul style="list-style-type: none"> <li>• 本科生获奖学金助学金总人数和总额</li> <li>• 全日制本科生获奖学金助学金总人数和总额</li> <li>• 本科生和研究生获服役奖学金(GI Bill)人数</li> </ul>		GI Bill: 指的是1944年的服役者的重新调整法案 ACT

\* 美国政府于1990年和2008年通过两个法令, 要求各大学必须向教育部和公众公布本校的毕业率。因此, 毕业率必须使用全国统一的计算方法, 其中同级组的定义是关键的一个变量。美国公立大学中, 一些全日制学生转为半日制, 所以同年入学的学生不一定同年毕业。可见, 同级组不等于同届毕业生。

美国高等教育数据库不仅储存资料, 更重要的是为使用者检索下载资料

服务。它提供了10种工具,以协助查询和下载所需要的统计报告和原始数据。1. 训练录像:使用者可以通过网站的录像,了解数据库的结构、定义、图表、数据收集表格,并解答共同性的问题。2. 趋势报告:使用者可以得到历年的学生入学人数,学位授予,毕业率,教职工信息,学校财政收入和学生资助信息。3. 学校资料:有两种方法,一种是索取所选择的大学的基本特点;另一种是得到所选择的大学的统计报告。4. 概要报告:各大学的校长可以得到每年一度本校的统计数据的概要报告。此报告以图表为主,概括主要指标。各校还可以选择同类学校,对照这些主要指标找到本校的差异。有两种选择方法:一种是自选对比组,另一种是遵循统一的标准,让数据库系统在全国范围内为使用者建立对比组。5. 统计报告:使用者可以选择任意一组学校的数据,做简单的统计分析,诸如总数、计数、平均数、中数、标准差和百分率。6. 专题报告:使用者可以选择任意一组学校,索取这些学校的专题信息,包括学杂费、住宿费、学生资助、招生标准、考试分数、在校生信息、学位授予和毕业率。7. 原始数据:数据库储存了七千多所高等院校和其他类型的学校的数据,共有近250个变量。使用者可以下载所需的原始数据(csv file)。8. 储存方式:1980—2005年的数据是压缩性资料(zip file)。2006年以后的数据储存在MSAccess。9. 快速查寻:常用此数据库的使用者可以保存以前用过的网页,并可以储存院校检索和变量检索的程序。10. 其他报告:包括介绍性报告、联网图表、数据采集方法论和摘要报告。

从美国高等教育数据库的结构看,有以下几个特点:第一,公开化。数据库的资料向公众开放,任何人都可以到网站查询资料和下载报告。使用者可以得到数据库工作人员的协助。第二,基础性。数据库采集六个领域的基本资料:在校生成简况、学校基本特点、招生标准、学位授予、学生持续率和毕业率、学费和学校资源。这些资料展示了各大学的基本概况,成为政策制定者和研究者们开展决策分析和高等教育研究的基础数据。第三,标准化。数据库建立了统一的数据采集标准,诸如全日制本科生和研究生等值的计算方法和毕业率的三个期限以及同级组的定义。这些标准和定义为数据采集的准确性和统一性奠定了基础。第四,操作性。该数据库与统计年鉴的根本区别是使用者可以运用资料进行各种分析,诸如同类院校的比较分析,5—10年的趋向分析,6个基本领域的专项分析等。

下面以呈交校长的概要报告中一个关于学杂费的比较资料图表,说明该数据库系统提供给学校的资料信息。图1中深色横条显示加州一所州立大学2013—2017年间4个学年的学杂费数据,浅色横条显示11所同类大学该指标的平均值。该校校长由此图可以看出本校在这4个学年度收取的学杂费与

同类高校的差异，以便做出下调或上调学杂费的决策，从而吸引到更多高质量生源，以及提高在校生的保留率。

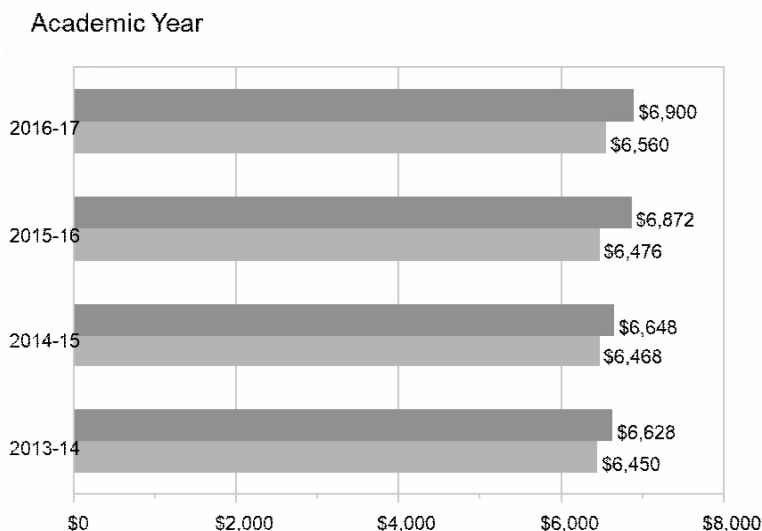


图1 某加州州立大学与该州其他11所同类大学学杂费的对比<sup>①</sup>

## 二、美国高等教育数据库的功能

美国高等教育数据库旨在为公众服务。这个大型数据库及其配备的10种工具，为不同的使用者提供不同的服务。简言之，它具有以下五项主要功能：

### (一) 为高校提供反馈和比较信息，改进学校管理

美国大学管理的一个重要环节是确保学生来源和新生质量。美国大学长期以来是供过于求，任何持有高中文凭的人只要交学费，都可以进入一所两年制的社区学院或四年制的大学。因此，大学招生市场上的竞争是比较激烈的。对于排名高的大学来说，主要是要争取优秀新生。为此，它们很注重同类大学的招生细则和录取分数，经常查寻这类数据，进行比较分析，以保持招生市场的优势。对于一般大学，特别是一般的私立院校来说，由于学费是学校主要的经费来源，确保充足的在校生数量则直接关系到学校的生存与发展。近年来，一些小型院校和女子学院，都由于生源不足而关门(Tarrant

<sup>①</sup> 资料来源：Figure 6. Academic year tuition and required fees for full-time, first-time degree/certificate-seeking undergraduates: 2013-14 to 2016-17, <https://www.csus.edu>。

et al, 2018)。因此,为了吸引学生,适当地调整学费成为必要的工具之一。为此,它们需要了解比较同类院校的学费,以确定最优价格,保持招生市场的已有份额。同时,各校学生获奖学金的比例也是吸引学生的重要指标。

美国大学管理的另一个环节是提高办学效益。大学是人力资本密集的机构,最大的开支是工资和福利。教学人员在总雇员中的比例,是衡量效益的一个指标。过多的行政人员会增加非教学性开支。在经费短缺的时期,最先裁减的是行政人员。从数据库中查询和比较同类院校的有关资料,及时调整教学人员在总雇员中的比例,是提高办学效益的重要措施。

从联邦教育部的角度看,数据库也为其制订政策提供了信息依据。例如,根据大学学费的增长,调节为低收入家庭提供的奖学金的数量。又如,注意到大学生群体构成的变化,增加了6年和8年毕业率为教育成果的指标。

## (二)为高校整体的评估认证提供重要数据

自1990年以来,多数认证协会从强调学校的“输入”,诸如教师质量,新生水平或教学设备,转变为强调学校的“输出”,即学生的学习经历和学习成果,并以此作为衡量学校高效绩的核心指标(Ewell, 1992)。

在数据库中8项有关学生的数据中,保持率、获取学位数和毕业率占4项。在10年一度的高校整体评估认证过程中,这些数据都是评估的重要指标(WASC 2013 Accreditation Handbook, Middle States 2015 Accreditation Standards)。认证小组不仅依据学校本身提供的统计数据,也参照美国高等教育数据库的资料(公布在各校的网站上)。如果学校提供的数据与高等教育数据库有出入,一般以数据库为准。

美国高校倡导多元化和教育平等,因而,学生群体的多元化,特别是少数民族在其总数的比例和他们的成就,也是评估认证的重点。众所周知,高等教育可以确立和改变获学位者的社会地位,特别是帮助低阶层群体向高阶层流动。长期以来,美国的黑人、西班牙裔和印第安人在大学生群体中占很小的比例,被称为“低入学率的少数民族”(underrepresented minority)。为了改变这种状况,许多大学采取各种措施,不仅增加“低入学率的少数民族”的入学率,而且尽力改进他们的保持率和毕业率。

与此相关的是教职工的多元化。女性和少数民族在其总数的比例,是评估认证的又一个指标。值得指出的是,教职工的少数民族包括亚裔人,因为他们在大学雇员中占的比例很低,与上述三个少数民族相似。倡导教职工的多元化不仅是遵循美国法律关于平等就业的条款,更重要的是少数民族教师对少数民族学生有不可取代的示范引导作用。

### (三)为高等教育的研究者提供基础资料

美国高等教育数据库是唯一的全国性资料来源。虽然它所储存的是基本数据,但许多重要的研究课题都是从这些数据中引发出来。例如,近年来,高校学费的快速增长成为公众注意的热点,也引起了高等教育研究者的关注。自2015年以来,发表在重要教育报刊杂志的有关高校学费的文章高达1447篇(平均一年近300篇)<sup>①</sup>。研究者正是从这个数据库中索取到5—10年全国大学的学费数据,进行趋向分析,发现了问题,并进一步寻找原因。

又如,随着高等教育的普及,高校毕业率成为教育研究的一个重要课题,称为“Access to Success”(Caruth, 2018)。从近年来的统计数字中,教育研究者发现,公立大学的4年毕业率很低。以美国最大的公立大学系统——加州州立大学系统为例,2015年,本科生4年毕业率为19%。经过采取各种措施,2018年毕业率才达到25%<sup>②</sup>。自2015年以来,教育研究者以该数据库的资料为基础,进行了多方面的探讨,诸如分析学生群体中低收入家庭半日制学生和少数民族学生的增长趋势(Glenn, 2017);对第一年新生保持率与4年毕业率做回归分析以及其他统计分析(Lenhardt, 2017),探讨改进毕业率的各种途径等(Romine et al., 2015)。这些研究都引用了高等教育数据库提供的各种基础数据信息。

### (四)为高等院校的排名榜提供主要数据

美国大学排名榜是由媒体主办的,种类很多。有以校园绿化和美化为主题的排名<sup>③</sup>,对校园伙食的排名,对学生时装的排名,更有甚者,以学生聚会为主题的快乐校园的排名<sup>④</sup>。其中最有权权威性的是美国《新闻与世界报道》主办的每年一度的大学排名榜。它包括所有的高等院校,将它们分为不同的类别,根据其制定的8项标准为各校打分,然后按分数进行质量排名。这8项标准是:第一年新生保持率和6年毕业率;低收入家庭学生的毕业率以及与其他学生毕业率的差距;6年毕业率的增长或下降;本科生的质量;教学人员的质量,包括工资福利,全日制比例,最高学位,教师/学生比例等;新生的接收率和新生的入学考试分数等;生均费用;校友捐款。

虽然该刊物每年向高校发放调查问卷和从学校网站的“通用资料”

① <https://www.eric.ed.gov>。

② <https://www.csus.edu>。

③ Tree Campus USA. <https://www.arborday.org>。

④ <https://www.princetonreview.com/college-rankings?rankings=best-385-colleges>。



(Common Data Set)采集信息。但是,每年都有相当一部分的大学对其问卷置之不理,也不一定及时更新他们网站的“通用资料”。特别是近年来,不少大学对美国新闻的排名标准提出异议,拒绝提供资料。面对这种局面,该刊物不得不依靠高等教育数据库。该刊物承认,在上面列出的8项标准中,除了校友捐款外,其余7项均来源于高等教育数据库<sup>①</sup>。

对各大学来说,排名榜是吸引学生,特别是吸引优秀生源的一个有效工具。同时,也对学校声誉有重大影响。很多院校对本校的年度排名密切注视。由于名次的变化,上移或下滑,都会激发有关院校采取一系列的改进措施。因此,高等教育数据库支持排名榜的信息采集,其目的是利用排名榜,推动各大学的改进。

#### (五)为高等院校的指南手册提供可靠数据

美国大学数量大种类繁多。根据2018年的统计,共有4324所四年和两年制大学,分为33种类型<sup>②</sup>。因此,学生和家長在选择学校时,需要指导和比较。许多大学指南手册编辑公司便应运而生。最初是每年出版和发行印刷版指南手册;现在则建立网站,发行电子版的指南手册。这类手册简要介绍各大学的基本情况和有关数据,并提供检索工具。

尽管排名榜和指南手册都是为学生和家長选择大学服务,但它们具有不同功能:排名榜只列出各校的8项分数和总分,而指南手册则提供高校办学条件、教育项目和学生资助等方面的细节信息。更重要的是,指南手册强调适宜性,旨在帮助学生找到最适合自己的大学,而不是盲目追求排名高的大学。以最著名的彼得森指南手册(Peterson's College Guidebook)为例,它提供以下信息:学校简介、地点和联系方法;学科和学位;招生规则、接受率和入学率、入学分数;学杂费;学生资助;学生群体构成以及多元化;校园生活和安全;体育运动;教学人员;附近的其他院校。

由此可见,以上多数信息可以直接从高等教育数据库索取。即是在排名榜和指南手册所依赖的各校网站提供的“通用资料”中,也要求其中许多数据必须与高等教育数据库的数据相符<sup>③</sup>。与此同时,宣称指南手册的数据来源于高等教育数据库,也为这些公司的出版物增加了可信性。

综上所述,联邦教育部的高等教育数据库不仅为其制定政策法规提供信息依据,而且通过以上五项功能,对高等教育的发展进行宏观调控。

① <https://www.usnews.com>。

② <https://www.carnegieclassifications.iu.edu>。

③ <https://www.petersons.com>。

### 三、美国高等教育数据库的建设条件及数据采集制度

建立高等教育数据库,需要一定的人力和物力资源。该数据库有五名工作人员,包括三名统计学家,一名教育研究人员和一名行政人员。他们的主要职责是指导信息采集过程,解答问题,帮助使用者检索数据和制作各种报告<sup>①</sup>。此外,数据输入、储存和输出则有专门设立的信息技术部门负责(Information Technology,简称IT)。IT为各大学建立密码保护的账户,各校设专人将索取的数据输入该网站,经检验后,将资料锁定。

全国性采集数据从每年从秋季学期开始,至春季学期结束。采集过程分三个阶段:秋季采集全年在校生总数,学位授予总数和获学位的学生人数和学校基本信息(#1, #4, #10);冬季采集招生标准和数据,毕业率和学生资助(#3, #7-8 #11-12);春季采集图书馆藏书,服务和预算;秋季新生入学人数;学校财政数据;教职工数据(#2, #5-6, #9)。每一阶段都规定呈交数据的最后限期。

为了确保资料的准确性,各高等院校几乎都设立了一个常设机构——学校管理研究处(Institutional Research,简称IR)。IR有多种功能,为高等教育数据库提供资料是它的一个主要职能。IR处长负有核查和锁定学校数据的责任。他/她不仅要建立发展本校特殊的数据库,为呈交主要资料做准备,还要与有关部门协作,包括图书馆、财务处和人事处等,共同采集所需要的其他有关资料。IR的成员多是持有学士或硕士学位的专业人员,不少IR处长持有博士学位。他们受过训练,懂数据库,熟悉高等教育数据库网上的资料输入和检索。他们既是数据提供者,更是数据的使用者。例如,IR的职责之一是为本校建立同类高校的比较组,每年下载数据库的概要报告,包括比较组高校的各类数据,呈送校长和其他领导成员。此外,许多州立大学系统(如加州州立大学系统)也设有IR,指定专人统筹本系统的资料采集,并由这个人核查和锁定系统内所有大学的资料。这一措施,有效地减少由不同人员采集信息的不一致性,增加了系统内大学可比数据资料的准确性。

鉴于资料的复杂性和变化性,美国联邦教育部每年都要举办培训班,既有定点授课,也有网上授课,帮助各校的IR人员了解数据库的要求、计算方法和新内容。在资料采集开始后,设有热线电话,为各校做及时指导和咨询。

高等教育数据库采集的数据有两个特点:第一个特点是采集冻结数据。

<sup>①</sup> <https://nces.ed.gov>。

各院校每学期都有一个固定的日期,由IT将本校数据库的有关信息复制下载,交由IR保存。经过清理和编排,此类冻结数据成为各校提交正式报告的基础。在拥有几万大学生的学校,数据每天都在变化,以非冻结数据为基础的报告缺乏连贯性、可比性。利用冻结数据是储存历史性资料 and 制作正式报告的唯一途径。第二个特点是采集聚合数据。各院校以冻结数据为基础,遵照高等教育数据库的要求,对数据进行加工,包括:建立同级组,然后追踪同级组4年,6年和8年的毕业率;计算全日制本科生和全日制研究生的当量;追踪同级组中享受联邦政府低收入助学金(Pell Grant)的学生的毕业率,等等。可见,各校提交的数据是经过加工的聚合数据,而不是原始的个人记录。而聚合数据必须由受过训练的IR专家来制作。

每年的数据采集结束后,数据库工作人员必需对数据进行核查,包括三个步骤:第一步是初步检查,发现不完全的或不合要求的资料,与这些院校联系以得到改正和补充;第二步是对资料进行综合,准备出版,并标明提供不完整资料的院校;第三步是让各院校进行最后的修改。一般只有1%—7%的数据会有变化。

如前所述,根据美国法律,联邦教育部有权对不提交或拖延提交资料的院校进行罚款,并给校长发警告信。因此,每年有7500多所高等院校和其他类型的高校向数据库提交资料,并按时锁定数据。如果因某种原因拖延,这些学校的IR必须加班完成。可见,法律程序保证了数据采集的高回收率和及时性。

#### 四、美国高等教育专题数据库

鉴于高等教育数据库只采集基本信息,远远不能满足许多高等教育管理者和研究者的需要。为此,以大学、协会或报刊牵头的全国性专题数据库不断涌现。本文只对其中几个大型专题数据库做简单介绍。

##### (一)高等院校教学人员工资和福利数据库(AAUP)

此数据库是由美国大学教授协会(American Association of University Professors,简称AAUP)建立的。AAUP成立于1915年。自1948年以来,该协会进行每年一度的问卷调查,采集关于高等院校教学人员工资和福利的资料,以及高级行政人员和作为教学助手的研究生的工资和福利数据。以此为基础,制作关于大学教学人员经济地位的年度报告。AAUP的资料采集遍布美国50个州,覆盖1000多所院校中37.5万名教学人员。更可观的是,在2018—2019年,AAUP共采集了952所院校中38万名全日制教学人员和360

所院校中 6.4 万名半日制教学人员的工资和福利的数据。

此数据库的经费来自协会成员的年度会费。会员院校提交资料后, 可以获取年度报告, 数据图表和原始资料。此外, 会员还可以建立对比组, 获取有关统计分析报告。到目前为止, AAUP 数据库一直是高校薪酬领域最权威性的信息来源<sup>①</sup>。

## (二) 大学生学习活动的问卷调查数据库(NSSE)

NSSE(National Survey of Student Engagement)数据库是由印第安纳大学主办的。1998年, 该大学酝酿了关于采集大学办学质量的问卷计划。1999年, 得到一个基金会的资助后, 开始了第一轮专对4年制大学学生的问卷。问卷采集一年级和四年级学生参与学校课内课外活动的信息, 以及这些活动对他们学习成就和个人成长的影响。这一问卷的新意在于将注意力从学生背景或分数转移到学生行为和活动, 诸如学生在学校的时间分配和自我评定的成长变化等。

参加年度问卷的学校要缴纳参与费, 提交抽样名单并选择发放问卷的日期。印第安纳大学设有问卷管理中心, 负责发放联网问卷, 催问电子邮件, 整理问卷结果和制作报告。与此同时, 各校的IR可以随时在网上观察问卷回收进程, 以便采取相应措施去增加回收率。鉴于各种原因, 此项问卷的平均回收率经常在24%左右。考虑到人力财力以及学生变化的渐进性, 许多大学基本上是每两年或每三年进行一次NSSE问卷, 因此, 每年大约有700—900多所4年制大学参加这一问卷活动。

NSSE中心为参与学校提供原始资料和综合报告。综合报告中包括各种复杂的统计分析和图表以及与同类学校和总体的比较资料。更有新意的是, 综合报告还提供了一年级和四年级学生之间的比较资料, 以此显示学生在接受4年大学教育后的变化。此外, 综合报告还从问卷中引伸出学生积极参与学校课内外活动的10个指标, 并总结出对学生成长最有影响的6项活动, 诸如新生学习组、学生参与教师的科研项目、海外留学等。这些报告和图表, 被各大学广泛地运用于改进教学, 准备学校整体评估, 加强教辅工作和课外活动等<sup>②</sup>。

## (三) 大学生保持率毕业率数据库(CSRDE)

CSRDE数据库(The Consortium for Student Retention Data Exchange,

<sup>①</sup> <https://www.aaup.org>。

<sup>②</sup> <http://nsse.indiana.edu>。

简称 CSRDE)<sup>①</sup>是由俄克拉荷马大学主办的。它是 1994 年由一位 IR 处长发起的,以采集 4 年制和 2 年制大学学生保持率和毕业率数据为主。1995 年发行了这个共同体的第一份统计报告。除了采集数据之外,CSRDE 还主持年度学术交流会,举办联网讲座和出版论文集。

到目前为止,共有 400 多所院校参加了这个共同体。参加者每年必须缴纳会费,以支持共同体的工作。与高等教育数据库相似,各院校提交冻结数据和聚合数据,包括本科生同级组人数,1—4 年的保持人数,4—6 年的毕业人数,分解为种族和性别。CSRDE 将提交的数据汇编成集,附加简单的统计分析,以电子版和印刷品的方式发放给共同体的大学。

这一数据库以及相关的活动,对研究如何改进本科生的保持率和毕业率提供了比较性资料,以帮助大学为本校的管理工作定位,并进一步寻找改进措施。

#### (四)高等院校教学成本核算和教师工作量数据库(The Delaware Cost Study)

这一数据库是由特拉华大学于 1992 创办的。至今有 200 多所大学为长期会员,先后有 600 多所院校参加。该数据库主要采集教学人员教学工作量、教学支出费用、科研费用等资料。由于不同学科的教学费用差异很大,例如音乐系、护理系、工程学院各系的教学支出比文学系历史系要高很多,分学科进行比较,是十分必要的。因此,这一数据库的主要特点是以学科为基本单位,进行跨校的横向比较。这一数据库得到了联邦教育部的高度评价。

此数据库着重于比较各学科的教学、科研和公益服务的支出,包括工资和流动资金(不含不动产的支出)。它以全美统一的学科定义(Classification of Instructional Programs,简称 CIP)为基础,运用系统化的精确的计算方法,帮助各校比较同类学科的费用,发现不合理的教学支出,支持学科和大学整体的评估认证,建立发展新学科和新项目。但是,这一数据库不适用于对学费或学校总支出的比较,也不能作为绩效拨款的依据<sup>②</sup>。

共同体的会员每年缴纳会费,并按照数据库的要求,提交冻结数据和经过计算的聚合数据。特拉华大学的数据库中心进行汇集和编制报告。采集、加工和汇编是一个非常复杂的过程,一般用一年的时间,才能将年度报告发给会员院校。

此外,还有一些自愿组成的小型数据库(Consortium),在此不一一介绍。

<sup>①</sup> <https://csrde.ou.edu>。

<sup>②</sup> <https://ire.udel.edu/cost>。

由大学指南公司和《美国新闻与世界报道》为排名榜设计的“通用资料”(Common Data Set),由报刊编制的高等教育统计年鉴(Almanac of Higher Education)等,也有一定的参考价值,但并非严格意义上的数据库。

总之,专题数据库一般由专业协会、大学或报刊主办,其经费均来自每年的会员费或参与费。会员或参与院校自愿参加,也可以随时退出。参加的院校负责提交数据,并可以查询和下载原始数据和综合报告。此类专题数据库是对美国高等教育数据库的有益补充,为各类院校制定改进教育教学等各项工作,制定发展规划和经费支出预算等提供了参考信息,是提升院校科学管理和决策水平的重要支撑。

## 五、对我国高等教育数据库建设的启示

有学者认为:“学科的发展有规律可循,一般都需要经历三个阶段:第一阶段是对基本事实和关键问题的发现,第二阶段是提出解释事实和问题的新理论,第三阶段是对新理论的检验、运用和预测。”(甘犁、冯帅章,2019)从国际高等教育顶级期刊少有中国学者对于中国高等教育问题的研究<sup>①</sup>,以及国内高校教育学科的国际排名(QS教育学科排名在前50的仅有北京师范大学教育学部)可以窥视,目前中国高等教育学科的发展尚处于第一阶段。这与我们国家尚没有形成高等教育管理和研究的基础性数据库不无关联。因为描述高等教育发展的基本状况(包括历史和当下),发现高等教育发展中存在的关键问题,都需要高质量、代表性强的数据库。缺乏这类基础数据库,成为我国高等教育研究进入国际前沿的桎梏,也成为我国高等教育学科向前发展的阻碍。

本文详细引介了美国联邦教育部建设的高等教育数据库(IPEDS)的结构和功能,以及负责该数据库运行的组织机构和数据采集制度,此外,还简要介绍了AAUP、NSSE等几项高等教育专题数据库的结构信息和运作方式,对我国高等教育数据库建设具有一定借鉴价值。

第一,高等教育数据库建设主体可以是官方和非官方,依据所需建设的数据库的性质来区分。官方,如教育部负责基础性综合性数据库的建设,类似于美国的IPEDS,这种数据库建设涉及高校数量多,范围广,包括各类本

---

<sup>①</sup> 2014—2018年中国学者在Higher Education, Assessment & Evaluation in Higher Education, Higher Education Policy上发表的文章数(142)占这三个期刊总发文量(1251)的11.35%。

专科院校，公办、民办高校。官方承办这类数据库具有行政和经费保障方面的优势。众所周知，开展这类全国范围高校数据采集工作，如由非官方机构承担，则需要巨额调查经费，而由官方，如教育部来承办，则教育部可以通过发布行政文件确保各类院校配合数据采集的工作，既节省了调查经费，也能够获得高质量、代表性强的数据。非官方机构负责或参与专题性数据库建设，这类数据库采集对象范围有限，采集的数据指标相对较少，比如 AAUP 仅需采集教师薪酬和福利方面的数据。由非官方机构来承办，具有专业性、灵活性等方面的优势。如美国印第安纳大学建设的 NSSE 数据库，NSSE 中心由高等教育研究专家负责设计大学生学习活动的调查问卷，同时，中心研究人员为参与学校提供原始资料和综合报告，综合报告中包括各种复杂的统计分析和图表以及与同类学校和总体的比较资料。因而，这类非官方的机构在数据库建设和服务方面具有专业性和灵活性。而且，这类数据库的运行方式可以采取会员制，解决了大型数据库建设的巨额资金问题。会员高校按照数据库建设章程缴纳会费，履行数据采集职责，同时享有免费获得相关数据和专题报告的权利。

第二，明确各类高等教育数据库的功能。开展数据库建设，首先需要明确建设的目的和功能，目的和功能确定下来，才可以着手数据库内容结构的设计，因为数据库内容结构是要服务于其功能的。由美国的几项大型高等教育数据库的功能可以发现，高等教育数据库的功能主要聚焦在四个方面，一是为大学整体的评估认证服务；二是为政府制定高等教育发展战略和相关政策服务；三是为高校提高办学质量、改进管理服务；四是为社会和公众关注的院校和学科排名服务。如 IPEDS 主要为联邦政府制订高等教育相关政策提供了信息依据；同时，也使得政府掌握了全国高等院校的发展状况和整体评估认证结果；此外，IPEDS 也成为美国《新闻与世界报道》等大学排行榜的主要数据来源。NSSE 数据库则为高校改进办学质量提供了专业性的分析报告。

第三，系统规划各类高等教育数据库的结构和指标体系。高等教育数据库的数据结构既要注重基础性，也要注重有延展性。因为数据库建设不是一蹴而就的，其基本框架在一定时期内相对稳定，而具体指标需要不断改进与丰富。此外，指标统计口径要统一化，数据采集方式要有可操作性。目前国内“由于各数据调查机构缺乏统一标准，造成条线整合难度较大。中央与地方的统计数据、非官方各机构公布的数据，因数据标准和统计口径不一，不仅较难整合，甚至互相矛盾，严重影响了数据的可信度。”（甘犁、冯帅章，2019）因此，我国高等教育数据库建设需要借鉴美国各类高等教育数据库建设的经验，强调数据结构的基础性、延展性，以及数据标准和统计口径的一致

性,从而提高来自不同数据库的数据可整合性,为更多复杂问题的高等教育研究和决策服务。

第四,建立数据库共享机制,最大程度地发挥各类高等教育数据库资源的社会效益。美国联邦教育部的 IPEDS 是供全美各类院校共享的数据库,而且国内外学者们也可以在网上下载需要的研究数据,最大程度地发挥了该数据库的社会效益。美国部分专题数据库虽然采取的是会员制,但非会员高校也可以在线浏览所发布的各类研究报告和所采用的数据采集工具,对于高等教育研究者也有很好的参考价值。国内现有的高等教育微观数据库建设还主要依赖于各高校学者或机构整合教育统计年鉴和高校科技统计资料汇编等,而且,目前由学者们和专业机构建设的高等教育数据库的公开透明度普遍较差,大多数只限于调查机构内部使用,缺乏数据共享机制。例如,第三方评估机构负责实施的全国本科教学评估和学科评估的基础数据库信息,尚没有公开,仅有他们发布的评估结果报告等资料。可喜的是,2017年由西南财经大学中国家庭金融调查与研究中心联合北京师范大学创新研究院等9个单位发起的中国高校数据调查共享平台就是针对目前国内社科界“数据孤岛”现象的一个有益的尝试。期待教育部和各专业机构打破数据壁垒,尽可能的让各类高等教育数据库之间实现有效整合、拼接,并建立数据库共享机制,为更多的学界人士开展高质量的研究,以及教育行政部门和高校管理者制定科学决策服务。

#### [参考文献]

- 甘犁、冯帅章,2019,《加强微观数据库建设,助推中国经济学发展》,《经济研究》第4期。
- Aliyeva, A. ,C. A. Cody, & K. Low, 2018, “The History and Origins of Survey Items for the Integrated Postsecondary Education Data System”, *National Postsecondary Education Cooperative Publications*, 23: 7-12.
- Caruth, G. D. , 2018, Student Engagement, Retention, and Motivation: Assessing Academic Success in Today's College Students”, *Participatory Educational Research*, 5(1): 17-30.
- Ewell, P. T. , 1992, Outcomes Assessment, Institutional Effectiveness, and Accreditation: A Conceptual Exploration, *The Resource Paper of Council on Postsecondary Accreditation Project*, pp. 7-11.
- Glenn, D. M. , 2017, “Persisting through College: The Academic and Social Integration of First-Generation College Students of Color Participating in a Student Engagement Program”, *Ed. D. Dissertation, University of St. Francis, ProQuest LLC*.
- Lenhardt, J. L. 2017, “A Comparative Trend Analysis of Institutional Variables and Retention Rates in Publicly Funded Institutions”, *Ph. D. Dissertation, Capella University*,



*ProQuest LLC.*

Romine, K. D. , R. M. Baker & K. A. Romine, 2018, “The Relationship between Administrative Intensity and Student Retention and Success: A Three-Year Study”, *Education Sciences*, v8 Article 159.

Tarrant, M. , N. Bray & S. Katsinas, 2018, “The Invisible Colleges Revisited: An Empirical Review”, *Journal of Higher Education*, 89(3): p341—367.

## Construction of American Higher Education Database and its Enlightenment

WANG Jing, HU Yong-mei

(1. Institutional Research, Sacramento State University;

2. Faculty of Education, Beijing Normal University)

**Abstract:** It is a long-term tradition and important mission for the Department of Education of United States to collect data and statistics. Integrated Post-secondary Education Data System(IPEDS) was established in 1986 and became a comprehensive higher education database since. This essay introduced the structure, functions and data collecting process of IPEDS. It also described several special databases with single subject, such as AAUP and NSSE, and their structures and operating process. To develop and enhance the higher education database in China, following could be considered: a) develop a comprehensive database by the government agency while developing several special databases by universities; b) define the categories and functions of each database clearly; c) set up the requirements and standardized calculations for data collection in order to obtain high quality data; and d) open all database systems to public to maximize their usage.

**Key words:** Integrated Post-secondary Education Data System(IPEDS); the structure of database; the function of database; the construction of database

(责任编辑: 刘泽云 责任校对: 刘泽云 孙志军)